

# Comparison of Primary Payer in Cancer Registry and Discharge Data

Youngmin Kwon, BA; Marcelo C. Perrailon, PhD; Coleman Drake, PhD; Bruce L. Jacobs, MD, MPH; Cathy J. Bradley, PhD; and Lindsay M. Sabik, PhD

State cancer registries capture most cancer diagnoses within a population and are an important data source for cancer surveillance and research.<sup>1</sup> One major body of research has relied on cancer registries to document and evaluate disparities among patients with cancer by types of insurance coverage, including comparing Medicaid beneficiaries with those covered by commercial insurance.<sup>2-9</sup> Many studies have examined disparities in cancer staging, treatment, and survival across insurance types based on the primary payer recorded in the registries.<sup>4,8,10-13</sup> They broadly conclude that insurance coverage and type of health plan, whether private or public or managed care vs fee-for-service, are associated with access to care and health outcomes.

However, the primary payer recorded in cancer registries may be imprecise, which could affect the estimates of disparities by types of coverage.<sup>14-17</sup> In the United States, complex coverage details of health insurance plans, such as public coverage provided through private managed care, may lead to misclassification of the primary payer. The potential for misclassification is particularly concerning when estimating disparities in cancer care associated with Medicaid, the largest single source of insurance in the United States, which provides coverage for low-income or disabled adults.<sup>18,19</sup> Because Medicaid primarily serves populations with pronounced barriers to care,<sup>20-23</sup> understanding the role and effectiveness of the program in addressing cancer-related disparities is critical. However, increasing and varying reliance on managed care organizations (MCOs) across states complicates accurate classification of Medicaid because many MCOs serve individuals enrolled in Medicaid, those eligible for Medicare, and those who purchase plans in private insurance markets.<sup>24</sup> Hence, there is an ample opportunity for miscoding across Medicaid, Medicare, and private insurance. The prospect of misclassification underscores the need to evaluate the payer variable in cancer registries to determine the appropriateness of registry-based analysis of disparities.

Therefore, our objective was to examine the primary payer variable in cancer registries using the Pennsylvania Cancer Registry (PCR) as a case study. We performed a linkage of the cancer registry data with hospital-based facility procedure records collected

## ABSTRACT

**OBJECTIVES:** To determine agreement between variables capturing the primary payer at cancer diagnosis across the Pennsylvania Cancer Registry (PCR) and statewide facility discharge records (Pennsylvania Health Care Cost Containment Council [PHC4]) for adults younger than 65 years, and to specifically examine factors associated with misclassification of Medicaid status in the registry given the role of managed care.

**STUDY DESIGN:** Cross-sectional analysis of the primary cancer cases among adults aged 21 to 64 years in the PCR from 2010 to 2016 linked to the PHC4 facility visit records.

**METHODS:** We assessed agreement of payer at diagnosis (Medicare, Medicaid, private, other, uninsured, unknown) across data sources, including positive predictive value (PPV) and sensitivity, using the PHC4 records as the gold standard. The probability of misclassifying Medicaid in registry was estimated using multivariate logit models.

**RESULTS:** Agreement of payers was high for private insurance (PPV, 89.7%; sensitivity, 83.6%), but there was misclassification and/or underreporting of Medicaid in the registry (PPV, 80%; sensitivity, 58%). Among cases with "other" and "unknown" insurance, 73.8% and 62.1%, respectively, had private insurance according to the PHC4 records. Medicaid managed care was associated with a statistically significant increase of 12.6 percentage points [95% CI, 9.4-15.8] in the probability of misclassifying Medicaid enrollment as private insurance in the registry.

**CONCLUSIONS:** Findings suggest caution in conducting and interpreting research using insurance variables in cancer registries.

*Am J Manag Care.* 2023;29(9):455-462. doi:10.37765/ajmc.2023.89425

## TAKEAWAY POINTS

Many researchers rely on cancer registries to study the impact of insurance coverage on cancer outcomes. However, complexities in coverage, including the use of private managed care plans within public programs, complicate precise classification of payers. Using Pennsylvania's cancer registry linked to state inpatient and outpatient facility visit records, we examined misclassification of payers. Our analysis revealed the following:

- ▶ Misclassification and underreporting of Medicaid insurance was high in cancer registry data.
- ▶ Medicaid managed care increases the probability that Medicaid is misclassified as private insurance.
- ▶ Researchers should be mindful of insurance data issues when conducting and interpreting analyses using payer information in registries.

## METHODS

## Data

We identified all first lifetime primary cancer cases among adults aged 21 to 64 years in the PCR from 2010 to 2016. The PCR is a population-based cancer registry that has been certified by the North American Association of Central Cancer Registries (NAACCR) and the National Program of Cancer Registries.<sup>27</sup> It includes information on patient demographics and details about incident cancer cases reported by multiple sources, including hospitals, doctor's offices,

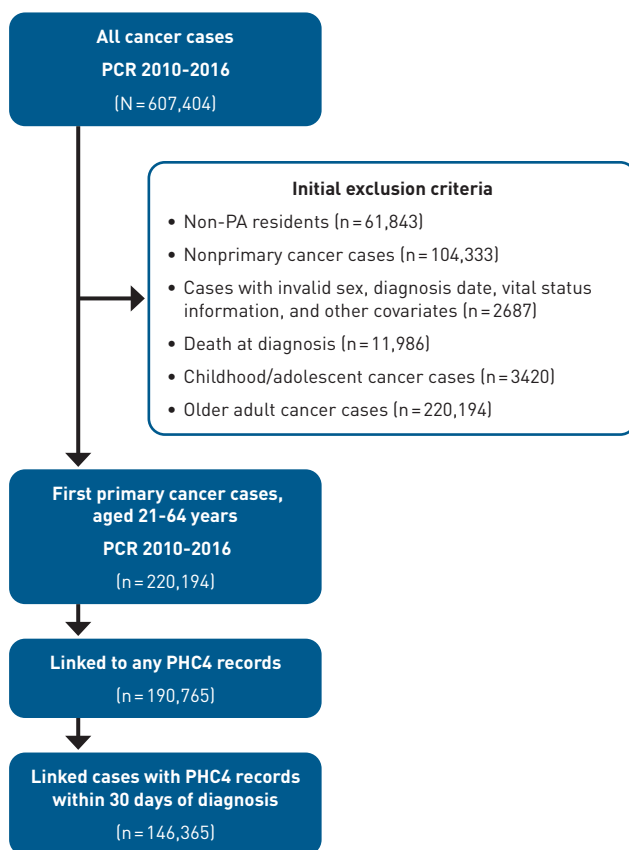
and death certificates. We excluded patients with missing or invalid diagnosis month information, a second cancer diagnosis within 365 days of the initial cancer incidence, or diagnosis date established through autopsy or death (Figure). The study was reviewed and determined exempt under Category 4 (secondary research) by the Institutional Review Board at the University of Pittsburgh.

Each case in the registry was linked to the PHC4 inpatient discharge and outpatient procedure records (2010-2017) using a unique masked patient identifier provided by state agencies. The PHC4 is an independent state agency authorized to collect public data from health care facilities in Pennsylvania.<sup>28</sup> It collects approximately 2 million inpatient hospital discharge and hospital outpatient procedure records every year. All licensed health care facilities in Pennsylvania are legally obligated to submit quarterly data that are validated before being made available to researchers. The inpatient records include all inpatient discharges, including psychiatric/behavioral health, rehabilitation, and drug and alcohol records (skilled nursing facility, swing bed, transitional care, 23-hour observation, and hospice records are excluded).<sup>29</sup> The outpatient records include a range of procedures performed within a hospital facility or other facilities associated with the hospital such as freestanding ambulatory surgery centers, ambulatory surgery rooms, short-term procedure units, laboratories, and the like.

Because the primary outcome of the analysis was insurance status around the time of diagnosis, all PHC4 records, rather than only cancer-related encounter records, were included. The linked data used for the current analysis also include information on time between the facility encounter and the date of cancer diagnosis. Any linked PHC4 records occurring more than 30 days before or after cancer diagnosis were excluded, given our focus on insurance status around the time of diagnosis.

Supplementary area-level data were merged with patient-level records based on patient Census tract of residence in the registry. Rural residence was defined based on rural-urban commuting area (RUCA) codes from the US Department of Agriculture, with urban areas defined as RUCA codes 1 to 3, large town areas as RUCA codes 4 to 6, and rural areas as RUCA codes 7 to 10.<sup>30</sup> The Area Deprivation Index (ADI) was used to capture Census tract-level socioeconomic

FIGURE. Inclusion/Exclusion Criteria Flow Chart, PCR 2010-2016



PA, Pennsylvania; PCR, Pennsylvania Cancer Registry; PHC4, Pennsylvania Health Care Cost Containment Council.

by an independent state agency (Pennsylvania Health Care Cost Containment Council, or PHC4<sup>25</sup>) and estimated the extent to which insurance variables agreed across the linked data sources. Further, we assessed factors associated with misclassification of Medicaid in the registry data, including enrollment in Medicaid MCOs, which cover almost 96% of Medicaid beneficiaries in Pennsylvania.<sup>26</sup>

status, with quartile 1 representing the least disadvantaged areas and quartile 4 representing the most disadvantaged.<sup>31</sup>

**Classification of Insurance**

Payer variable coding and categories differ across the PCR and the PHC4 records. In the registry, a single “primary payer at diagnosis” variable records the patient’s primary insurance at the time of diagnosis and/or treatment,<sup>32</sup> whereas the PHC4 records contain 3 payer variables, populated based on all sources of payment for each visit. We created a mutually exclusive insurance variable that summarizes payer information from each data source using 6 categories in the following hierarchy: Medicare (including Medicare-Medicaid duals), Medicaid, private, other, uninsured, and unknown. For the registry data, we recoded the predefined categories; for the PHC4 records, we used a combination of the 3 payer variables to designate insurance status (eAppendix A [eAppendices available at [ajmc.com](http://ajmc.com)]). When classifying insurance, we used the PHC4 record closest to the diagnosis within the 30-day window around the time of diagnosis.

**Statistical Analysis**

We performed a cross-tabulation of insurance variables across the 2 data sources to calculate the positive predictive value (PPV) and sensitivity of the payer information in the registry, using the PHC4 as the gold standard based on the assumption that the PHC4 records are more precise because they are based on billing information from a single inpatient or outpatient visit and are recorded in a more detailed manner for reimbursement purposes. In the context of our study, PPV measures how well the registry accurately captures a pool of patients within a particular payer category. On the other hand, sensitivity measures whether the registry correctly categorized the payer sources as listed on the PHC4 records. In addition, we calculated Cohen  $\kappa$  statistics to test the reliability of insurance categorization between the 2 data sources.

To predict the likelihood of misclassifying Medicaid in the registry, we limited the sample to cases classified as having Medicaid (either managed care or fee-for-service) in the PHC4 records and estimated logit models with Medicaid misclassifications in the registry (overall misclassification and misclassification as private insurance) as outcomes while controlling for managed care enrollment and other patient- and area-level covariates. We clustered the SEs at the PHC4 facility level because the classification of payers is likely correlated with coding practices within each facility. We report the marginal effects of misclassification. All analyses used SAS version 9.0 (SAS Institute) and Stata/MP 17 (StataCorp).

We also conducted 2 stratified analyses. First, we stratified our main analysis to years before (2010-2014) and after (2015-2016) Pennsylvania expanded Medicaid to cover newly eligible adults with incomes below 138% of the federal poverty line<sup>33</sup> because Medicaid expansion changed the composition of the population covered by Medicaid<sup>34</sup> and increased enrollment in managed care plans to accommodate the newly eligible enrollees.<sup>35</sup> Second, we

**TABLE 1.** Characteristics of First Primary Cancer Cases in PCR With Linked PHC4 Records, 2010-2016

Variable	Full analysis sample	Medicaid analysis sample
	n = 146,365	n = 23,160
Sex, n (%)		
Female	84,226 (57.5%)	12,843 (55.5%)
Male	62,139 (42.5%)	10,317 (44.5%)
Age in years		
Mean (SD)	52.7 (9.4)	50.9 (10.2)
Median (IQR)	55.0 (48.0-60.0)	53.0 (46.0-59.0)
Range	21.0-64.0	21.0-64.0
Age group in years, n (%)		
< 45	25,493 (17.4%)	5248 (22.7%)
45-49	17,104 (11.7%)	2949 (12.7%)
50-54	27,700 (18.9%)	4647 (20.1%)
55-59	35,887 (24.5%)	5299 (22.9%)
60-64	40,181 (27.5%)	5017 (21.7%)
Race/ethnicity, n (%)		
Hispanic	3525 (2.4%)	1465 (6.3%)
Non-Hispanic White	117,743 (80.4%)	14,826 (64.0%)
Non-Hispanic Black	15,573 (10.6%)	5575 (24.1%)
Non-Hispanic other	2571 (1.8%)	530 (2.3%)
Unknown	6953 (4.8%)	764 (3.3%)

(continued)

separately analyzed misclassifications by 2 age groups—adults younger vs older than 50 years—to account for higher incidence of cancer among older adults.<sup>36</sup>

**RESULTS**

Nearly all cases in the registry had at least 1 PHC4 record (n = 190,765; 94%). Our analytic sample consisted of the 77% (n = 146,365) of matched cases in the registry that had a linked record within 1 month before or after cancer diagnosis. A little more than half of the analytic sample also had a record coinciding with the day of cancer diagnosis (n = 81,693; 56%). The patients included in the analytic sample tended to be women (57.5%), older than 45 years (82.6%), and non-Hispanic White (80.4%); have localized cancer (40.7%); and live in urban areas (87.1%) (Table 1). Compared with the full sample, there was a larger proportion of non-Hispanic Black patients (24.1%), lung cancer (16.2%), distant cancer stage (33.8%), and individuals in the fourth-quartile ADI group (49.5%) among cases classified as having Medicaid in the PHC4 records. In the full analytic sample, the most common source of coverage was private insurance (64.6%), followed by Medicaid (11.5%).

The comparison of the payers revealed varying levels of agreement between the registry and the PHC4 records by insurance type (Table 2). PPV was moderately high for Medicaid (80%), meaning that 80% of patients for whom the primary payer in the registry was

**TABLE 1.** (Continued) Characteristics of First Primary Cancer Cases in PCR With Linked PHC4 Records, 2010-2016

Variable	Full analysis sample	Medicaid analysis sample
	n = 146,365	n = 23,160
Cancer type, n (%)		
Lung	15,380 (10.5%)	3745 (16.2%)
Breast	30,342 (20.7%)	3293 (14.2%)
Prostate	8013 (5.5%)	800 (3.5%)
Colorectal	14,130 (9.7%)	2319 (10.0%)
Other	78,500 (53.6%)	13,003 (56.1%)
Cancer stage, n (%)		
In situ	14,054 (9.6%)	1462 (6.3%)
Localized	59,636 (40.7%)	7344 (31.7%)
Regional	33,273 (22.7%)	5652 (24.4%)
Distant	35,435 (24.2%)	7831 (33.8%)
Unknown	3967 (2.7%)	871 (3.8%)
Year of diagnosis, n (%)		
2010	20,434 (14.0%)	3003 (13.0%)
2011	20,691 (14.1%)	3271 (14.1%)
2012	20,493 (14.0%)	3216 (13.9%)
2013	20,729 (14.2%)	3195 (13.8%)
2014	21,021 (14.4%)	3098 (13.4%)
2015	21,559 (14.7%)	3497 (15.1%)
2016	21,438 (14.6%)	3880 (16.8%)
Area Deprivation Index		
Mean (SD)	50.5 (23.5)	64.1 (22.4)
Median (IQR)	50.0 (32.5-69.0)	67.8 (47.2-83.3)
Range	1.0-100.0	1.0-100.0
Quartile 1, n (%)	36,990 (25.3%)	2436 (10.5%)
Quartile 2, n (%)	35,713 (24.4%)	4002 (17.3%)
Quartile 3, n (%)	35,693 (24.4%)	5264 (22.7%)
Quartile 4, n (%)	37,969 (25.9%)	11,458 (49.5%)
Region, n (%)		
Urban	127,505 (87.1%)	19,977 (86.3%)
Large town	12,406 (8.5%)	2065 (8.9%)
Small town	3569 (2.4%)	623 (2.7%)
Rural	2884 (2.0%)	495 (2.1%)
PCR primary payer at diagnosis		
Uninsured	2847 (1.9%)	N/A
Medicaid	16,776 (11.5%)	N/A
Medicare	14,284 (9.8%)	N/A
Private	94,611 (64.6%)	N/A
Other	15,845 (10.8%)	N/A
Unknown	2002 (1.4%)	N/A

N/A, not applicable; PCR, Pennsylvania Cancer Registry; PHC4, Pennsylvania Health Care Cost Containment Council.

Medicaid indeed were covered by Medicaid according to the PHC4 records. However, sensitivity was low (58%), indicating that among Medicaid-insured individuals in the PHC4 records, the registry only classified approximately half of them as having Medicaid. In contrast, both PPV and sensitivity were high for private insurance (89.7% and 83.6%, respectively). In addition, the rate of misclassification was high for patients who had “other” or “unknown” insurance, most of whom had private insurance according to the PHC4 records. Finally, 42.7% of patients classified as uninsured in the registry were insured by Medicaid according to the PHC4 records. The  $\kappa$  statistic was 0.544 (95% CI, 0.540-0.547), indicating a moderate level of agreement in insurance classifications.

In the logit models (Table 3), having Medicaid managed care was associated with a higher probability of overall misclassification of Medicaid (across any other insurance category) and misclassification of Medicaid as private insurance by 2.6 percentage points (95% CI, -2.1 to 7.3;  $P = .278$ ) and 12.6 percentage points (95% CI, 9.4-15.8;  $P < .001$ ), respectively. The likelihood of any Medicaid misclassification was lower among non-Hispanic Black patients compared with non-Hispanic White patients and among those living in nonurban areas or the highest ADI quartile. The likelihood of Medicaid misclassification as private insurance specifically was lower among Hispanic patients and all age groups older than 50 years but higher among those who lived in the lowest ADI quartile. The likelihood of misclassification as private insurance remained significant and consistent when stratified by age groups and time periods before and after Medicaid expansion (Table 4, eAppendix B, and eAppendix C).

## DISCUSSION

We examined the concordance of the primary payer variable in a large statewide cancer registry by linking the PCR with Pennsylvania's facility encounter records data, contributing to the limited prior literature that explored linkages of the registries with other data sources such as Medicaid and Medicare claims and enrollment files,<sup>15,37,38</sup> all-payer claims databases (APCDs),<sup>39,40</sup> hospital registries,<sup>41</sup> and electronic health records claims from health systems.<sup>42,43</sup> In our study, we found only a moderate level of agreement in insurance classification between the registry and the facility encounter records ( $\kappa = 0.544$ ). Although the agreement was high for private insurance, there were substantial misclassification and underreporting of Medicaid in the registry. Moreover, a sizeable proportion of cases in the registry with unknown/other insurance was identified as having private insurance in the PHC4 records. Most patients who were classified as uninsured in the registry had Medicaid according to the PHC4 records. Lastly, Medicaid managed care was significantly associated with a higher probability of misclassifying Medicaid as private insurance. Overall, these findings demonstrate the limitations of the primary payer variable in cancer registries in ascertaining sources of insurance coverage among adult patients with cancer who are younger than 65 years.

**TABLE 2.** Payer Agreement and Misclassification Across PCR and PHC4 Records Measured by PPV and Sensitivity<sup>a</sup>

		PHC4 insurance within a month of diagnosis					
		Uninsured (n = 2859; 2.0%)	Medicaid (n = 23,160; 15.8%)	Medicare (n = 15,635; 10.7%)	Private (n = 101,565; 69.4%)	Other (n = 2446; 1.7%)	Unknown (n = 700; 0.5%)
Uninsured (n = 2847; 1.9%)	Row % (column %)	37.4 (37.3)	42.7 (5.2)	1.7 (0.3)	15.7 (0.4)	0.7 (0.9)	1.8 (7.3)
Medicaid (n = 16,776; 11.5%)	Row % (column %)	4.8 (27.9)	80.0 (58.0)	6.7 (7.2)	7.1 (1.2)	0.9 (6.2)	0.5 (12.3)
Medicare (n = 14,284; 9.8%)	Row % (column %)	0.7 (3.4)	7.2 (4.5)	76.6 (70.0)	14.7 (2.1)	0.6 (3.4)	0.2 (4.6)
Private (n = 94,611; 64.6%)	Row % (column %)	0.5 (16.3)	5.4 (22.1)	3.0 (18.0)	89.7 (83.6)	0.9 (36.3)	0.4 (58.6)
Other (n = 15,845; 10.8%)	Row % (column %)	1.9 (10.7)	12.4 (8.4)	3.2 (3.2)	73.8 (11.5)	8.1 (52.2)	0.7 (15.4)
Unknown (n = 2002; 1.4%)	Row % (column %)	6.3 (4.4)	19.9 (1.7)	9.7 (1.2)	62.1 (1.2)	1.3 (1.1)	0.6 (1.9)

PCR, Pennsylvania Cancer Registry; PHC4, Pennsylvania Health Care Cost Containment Council; PPV, positive predictive value.

<sup>a</sup>Each cell includes a row percentage and a column percentage (in parentheses). The darker shaded cells along the diagonal indicate the agreement across the 2 data sources, and the unshaded cells indicate misclassification. Row percentages in the shaded cells are interpreted as PPV, or the proportion of patients within each PCR payer category that had a matching payer in the PHC4 records [gold standard]. Column percentages in the shaded cells are interpreted as sensitivity, or the proportion of patients within each PHC4 payer category that had a matching payer in PCR. The "other" category includes TRICARE, veterans' benefits, Indian Health Service, or insurance that is not otherwise specified. The "private" category includes all nonpublic plans, including marketplace plans, employer-sponsored insurance, Blue Cross, or other commercial plans.

Our results are comparable to those of earlier studies using registry/administrative data linkages. For example, Chan et al used the Medi-Cal enrollment files and reported that there was a moderate PPV (77%) and a poor sensitivity (48%) for Medicaid in the California cancer registry.<sup>15</sup> Perrailon et al linked the Colorado Central Cancer Registry to the state's APCD and found a higher PPV and sensitivity for Medicaid (97% and 70%, respectively) and a similar agreement for private insurance (86%-88% for both PPV and sensitivity) (unpublished data, October 2022). The higher rate of agreement in their study may result from the inclusion of all claims (as opposed to only facility-based records) and the fact that Colorado's Medicaid program does not have managed care plans administered by private insurance companies. In contrast, Medicaid MCOs covered nearly all Medicaid enrollees in Pennsylvania in 2021.<sup>26</sup>

The primary payer variable in cancer registries has been recognized by registrars as one of the most difficult fields to populate due to lack of uniform training and the complex and constantly evolving landscape of insurance plans.<sup>16,44</sup> This could explain the finding that enrollment in a Medicaid MCO significantly increased the probability of misclassification as private insurance, even though the reporting manual for the PCR specifies a category for Medicaid administered through an MCO.<sup>32</sup> Because MCOs often serve enrollees across Medicaid, Medicare, and private insurance plans, there is a greater potential for miscoding public insurance such as Medicaid as private insurance. We also found that the association between Medicaid managed care and misclassification in the registry remained significant in years after Pennsylvania's Medicaid expansion, suggesting that misclassification will persist with increasing penetration of Medicaid managed care.

Given the evidence of misclassification, we suggest caution in using and interpreting analyses that rely on the payer variable in cancer registry data. Although we do not formally quantify the impacts of such misclassifications on the estimates of disparities, our findings behoove researchers to consider the appropriateness of

the registry-based payer variable in specific contexts. For example, the use of the payer variable may be reasonable when defining a sample of cases within each payer category, especially for payers such as private insurance, which had a high PPV. However, comparing outcomes between payers can be problematic due to the low sensitivity for certain payer categories. Commonly, studies of disparities in cancer outcomes compare Medicaid beneficiaries with privately insured individuals. If we simply assume that a substantial portion of patients with Medicaid is mostly misclassified as having private insurance (42% in our study), then the comparison will bias the estimates of disparities toward the null due to the cross-sorting of cancer cases between Medicaid and private insurance. However, there may be miscoding of Medicaid and private insurance as other categories as well. This, combined with selection into plans of different benefit designs and care quality, suggests that bias could run in either direction that may not be predicted ex ante. We conducted a simple, unadjusted assessment of disparities in 5-year mortality rates between Medicaid and privately insured patients diagnosed with 4 major cancers (breast, colorectal, lung, and prostate) using the payer variable in the registry vs the PHC4 records and found that the discordance in the registry generally underestimated the estimates of disparities across cancer types and stages. This is consistent with our hypothesis that misclassification across these 2 payer types will attenuate disparities estimates (eAppendix D).

Although our study of misclassifications is specific to the use of cancer registries, it represents an important case study that illuminates one of many data-related issues that researchers may encounter when examining the highly complex system of insurance coverage in the United States. In particular, as the boundary between public and private insurance becomes increasingly blurred, researchers must grapple with complicated questions about how to conceptualize different insurance classifications.<sup>45</sup> For individuals enrolled in private managed care plans under Medicaid, the primary payer is Medicaid, but depending on the research question of

**TABLE 3.** Marginal Effects From Logit Models Estimating Probabilities of Misclassification of Medicaid in Pennsylvania Cancer Registry<sup>a</sup>

	Any misclassification			Misclassification as private		
	Marginal effects	95% CI	P	Marginal effects	95% CI	P
Medicaid delivery system (ref: fee-for-service)						
Managed care	2.6	-2.1 to 7.3	.278	12.6***	9.4-15.8	<.001
Sex (ref: female)						
Male	0.2	-1.8 to 2.3	.838	-1.3	-2.9 to 0.2	.088
Age in years (ref: <45)						
45-49	0.0394	-2.4 to 2.4	.974	-1.3	-3.3 to 0.7	.212
50-54	-0.2	-2.1 to 1.6	.795	-1.9*	-3.6 to -0.3	.023
55-59	0.0058	-1.5 to 1.5	.994	-2.7***	-4.2 to -1.3	<.001
60-64	3.9***	2.0-5.9	<.001	-3.3***	-4.7 to -1.8	<.001
Race/ethnicity (ref: non-Hispanic White)						
Hispanic	-5.3	-12.5 to 2.0	.158	-8.8***	-11.6 to -5.9	<.001
Non-Hispanic Black	-7.6**	-13.1 to -2.1	.007	-2.6	-6.0 to 0.8	.134
Non-Hispanic other	-4.6	-11.3 to 2.2	.185	0.3	-4.0 to 4.5	.898
Unknown	-2.5	-11.3 to 6.4	.585	-3.8	-10.7 to 3.1	.284
Cancer type (ref: other)						
Lung	1.9	-0.6 to 4.4	.134	0.4	-1.5 to 2.3	.676
Breast	3.3*	0.1-6.5	.046	1.9	-0.8 to 4.5	.165
Prostate	4.7*	0.7-8.8	.022	2.2	-1.2 to 5.5	.200
Colorectal	2.8*	0.2-5.3	.033	2.7**	0.8-4.6	.006
Cancer stage (ref: localized)						
In situ	-1.5	-5.7 to 2.8	.505	-2.1	-4.9 to 0.7	.149
Regional	-1.3	-3.1 to 0.4	.135	0.0154	-1.2 to 1.2	.98
Distant	-1.7	-4.1 to 0.7	.164	-0.6	-2.2 to 0.9	.432
Unknown	1.7	-2.9 to 6.3	.470	-3.8**	-6.7 to -1.0	.009
Region (ref: urban)						
Large town	-5.8**	-10.1 to -1.4	.009	-0.6	-4.1 to 2.9	.725
Small town	-9.4**	-15.2 to -3.6	.001	-1.4	-6.3 to 3.6	.586
Rural	-4.2	-10.3 to 1.9	.179	-0.7	-7.0 to 5.6	.818
ADI (ref: quartile 4)						
Quartile 1	11.8***	7.4-16.2	<.001	8.9***	6.2-11.6	<.001
Quartile 2	10.2***	6.0-14.5	<.001	2.3	0.0-4.7	.050
Quartile 3	3.7**	1.1-6.2	.005	1.8	-0.2 to 3.8	.086
Year of diagnosis (ref: 2016)						
2010	2.7	-3.2 to 8.7	.366	0.3	-3.8 to 4.4	.880
2011	3.4	-1.3 to 8.0	.157	0.9	-2.8 to 4.6	.637
2012	0.4	-4.5 to 5.3	.884	-2.3	-5.7 to 1.2	.197
2013	5.1*	0.5-9.7	.028	-0.4	-3.6 to 2.9	.830
2014	5.4*	0.9-9.9	.018	0.3	-3.3 to 3.8	.876
2015	1.3	-1.9 to 4.6	.424	-0.4	-3.0 to 2.3	.789

ADI, Area Deprivation Index; ref, reference.

\**P* < .05; \*\**P* < .01; \*\*\**P* < .001.

<sup>a</sup>The marginal effects from the logit models were multiplied by 100 to represent percentage-point changes in misclassification probabilities. The SEs are clustered at the Pennsylvania Health Care Cost Containment Council facility level.

interest, researchers need access to high-quality data that accurately and consistently account for granular coverage details, such as private administration of public coverage. We demonstrate how linkages to ancillary data sources (such as PHC4) that contain multiple, detailed payer-type variables may support more nuanced analyses of coverage.<sup>46</sup> Such linkages may be also valuable in investigating secondary insurance, which is relevant for adults older than 65 years, most of whom have Medicare supplemental coverage. For example, multiple payer-type variables may better facilitate identification of individuals who are dually eligible for Medicare and Medicaid who may be coded as having either Medicaid or Medicare (depending on how one conceptualizes the primary payer for these individuals) but not “Medicare with Medicaid eligibility,” which is one of the categories for the primary payer variable in the NAACCR system.<sup>47</sup>

**Limitations**

We acknowledge that the facilities data used in our study are provided by hospitals and outpatient facilities and do not represent a true gold standard in verifying the insurance information. However, we assume that they are more reliable sources to ascertain coverage information because they are generated by facilities that have financial incentives to accurately and comprehensively code for reimbursement purposes.<sup>48</sup> Our linked data also include most, if not all, hospital-based records for each cancer case, which is an improvement from registry-only data that record insurance information at a single point in time. Moreover, compared with other studies using Medicaid enrollment records or APCDs, our data also capture the uninsured, which is an important population to monitor because they are more likely to experience suboptimal cancer outcomes compared with the insured.<sup>7</sup> Even so, facility records are not perfect. Retroactive coverage,<sup>49</sup> for example, is expected to be reflected in the PHC4 records because hospitals report the payer from which they expect to receive reimbursement, but we cannot rule out the possibility that a portion of uninsured and Medicaid-insured patients is misclassified because the data do not capture the complexity of presumptive eligibility or retroactive enrollment.

## CONCLUSIONS

Considering the high rate of misclassification in cancer registries, we recommend that researchers supplement the registry with ancillary data sources whenever possible for analyses that require detailed and accurate insurance coverage information. At a minimum, researchers should consider how the unique insurance landscape (including managed care penetration) in each state may create state-level variation in payers and populations covered by different payers when conducting and interpreting such analyses. ■

**Author Affiliations:** Department of Health Policy and Management, University of Pittsburgh School of Public Health (YK, CD, LMS), Pittsburgh, PA; Department of Health Systems, Management & Policy, University of Colorado Anschutz Medical Campus (MCP, CJB), Aurora, CO; Department of Urology, Division of Health Services Research, University of Pittsburgh School of Medicine (BLJ), Pittsburgh, PA; Colorado School of Public Health (CJB), Aurora, CO.

**Source of Funding:** The study was funded by a grant from the Agency for Healthcare Research and Quality (R01HS027396).

**Author Disclosures:** Mr Kwon and Drs Drake and Sabik report grant funding from the Agency for Healthcare Research and Quality (R01HS027396). The remaining authors report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

**Authorship Information:** Concept and design (MCP, BLJ, LMS); acquisition of data (CJB, LMS); analysis and interpretation of data (YK, MCP, CD, BLJ, CJB, LMS); drafting of the manuscript (YK, MCP, LMS); critical revision of the manuscript for important intellectual content (YK, MCP, CD, BLJ, CJB, LMS); statistical analysis (YK, MCP, CD, LMS); obtaining funding (LMS); and supervision (BLJ, LMS).

**Address Correspondence to:** Youngmin Kwon, BA, University of Pittsburgh School of Public Health, A610 Public Health, 130 DeSoto St, Pittsburgh, PA 15261. Email: yok84@pitt.edu.

## REFERENCES

- 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. doi:10.1002/cncr.30905
- Bradley CJ, Gardiner J, Given CW, Roberts C. Cancer, Medicaid enrollment, and survival disparities. *Cancer*. 2005;103(8):1712-1718. doi:10.1002/cncr.20954
- Kneuzert PJ, Kao LS, Ko TC, Wray CJ. Regional disparities affect treatment and survival of patients with intrahepatic cholangiocarcinoma—a Texas Cancer Registry analysis. *J Surg Oncol*. 2014;110(4):416-421. doi:10.1002/jso.23664
- Sung JM, Martin JW, Jefferson FA, et al. Racial and socioeconomic disparities in bladder cancer survival: analysis of the California Cancer Registry. *Clin Genitourin Cancer*. 2019;17(5):e995-e1002. doi:10.1016/j.clgc.2019.05.008
- Walker GV, Grant SR, Guadagnolo BA, et al. Disparities in stage at diagnosis, treatment, and survival in nonelderly adult patients with cancer according to insurance status. *J Clin Oncol*. 2014;32(28):3118-3125. doi:10.1200/JCO.2014.55.6258
- Ellis L, Canchola AJ, Spiegel D, Ladabaum U, Haite R, Gomez SL. Trends in cancer survival by health insurance status in California from 1997 to 2014. *JAMA Oncol*. 2018;4(3):317-323. doi:10.1001/jamaonc.2017.3846
- Niu X, Roche LM, Pawlish KS, Henry KA. Cancer survival disparities by health insurance status. *Cancer Med*. 2013;2(3):403-411. doi:10.1002/cam4.84
- Coburn N, Fulton J, Pearlman DN, Law C, DiPaolo B, Cady B. Treatment variation by insurance status for breast cancer patients. *Breast J*. 2008;14(2):128-134. doi:10.1111/j.1524-4741.2007.00542.x
- Roetzheim RG, Pal N, Gonzalez EC, Ferrante JM, Van Durme DJ, Krischer JP. Effects of health insurance and race on colorectal cancer treatments and outcomes. *Am J Public Health*. 2000;90(11):1746-1754. doi:10.2105/ajph.90.11.1746
- Berrian JL, Liu Y, Lian M, Schmalz CL, Colditz GA. Relationship between insurance status and outcomes for patients with breast cancer in Missouri. *Cancer*. 2021;127(6):931-937. doi:10.1002/cncr.33330
- Parikh AA, Robinson J, Zaydfudim VM, Penson D, Whiteside MA. The effect of health insurance status on the treatment and outcomes of patients with colorectal cancer. *J Surg Oncol*. 2014;110(3):227-232. doi:10.1002/jso.23627
- Parikh-Patel A, Morris CR, Kizer KW. Disparities in quality of cancer care: the role of health insurance and population demographics. *Medicine (Baltimore)*. 2017;96(50):e9125. doi:10.1097/MD.00000000000009125
- Shah AA, Sun Z, Eom KY, et al. Treatment disparities in muscle-invasive bladder cancer: evidence from a large statewide cancer registry. *Urol Oncol*. 2022;40(4):164.e17-164.e23. doi:10.1016/j.urolonc.2021.12.004
- Sabik LM, Bradley CJ. Understanding the limitations of cancer registry insurance data—implications for policy. *JAMA Oncol*. 2018;4(10):1432-1433. doi:10.1001/jamaonc.2018.2436

**TABLE 4.** Marginal Effects From Logit Models Estimating Probabilities of Misclassification of Medicaid in Pennsylvania Cancer Registry, Stratified by Age Groups and Years Before and After Medicaid Expansion\*

Stratification	Any misclassification			Misclassification as private		
	Marginal effects	95% CI	P	Marginal effects	95% CI	P
Overall	2.6	-2.1 to 7.3	.278	12.6***	9.4-15.8	<.001
Aged <50 years	3.2	-1.4 to 7.9	.173	13.1***	9.8-16.4	<.001
Aged ≥50 years	1.5	-4.0 to 6.9	.599	11.7***	8.1-15.3	<.001
Pre-Medicaid expansion (2010-2014)	3.6	-1.3 to 8.5	.146	14.0***	10.6-17.4	<.001
Post Medicaid expansion (2015-2016)	-0.6	-6.6 to 5.4	.841	8.7***	4.8-12.6	<.001

\*\*\*P < .001.

\*The marginal effects shown are the coefficients for the Medicaid managed care variable from Table 3. They were multiplied by 100 to represent percentage-point changes in misclassification probabilities. The SEs are clustered at the Pennsylvania Health Care Cost Containment Council facility level.

- Chan JK, Gomez SL, O'Malley CD, Perkins CI, Clarke CA. Validity of cancer registry Medicaid status against enrollment files: implications for population-based studies of cancer outcomes. *Med Care*. 2006;44(10):952-955. doi:10.1097/01.mlr.0000220830.46929.43
- Sherman RL, Williamson L, Andrews P, Kahn A. Primary payer at DX: issues with collection and assessment of data quality. *J Registry Manag*. 2016;43(2):99-100.
- Gershman S, Weiss N, Knowlton R, Solis A, Das B. An assessment of the primary payer variable among breast and colorectal cancer cases in the Massachusetts Cancer Registry, 2005-2009. *J Registry Manag*. 2017;44(4):143-145.
- Rudowitz R, Garfield R, Hinton E. 10 things to know about Medicaid: setting the facts straight. Kaiser Family Foundation. March 6, 2019. Accessed February 25, 2022. <https://www.kff.org/web/20220210161030/>
- Gabow P, Daschle T. Fifty years later: why Medicaid still matters. *Health Affairs*. July 29, 2015. Accessed February 25, 2022. <https://www.healthaffairs.org/doi/10.1377/forefront.20150729.049632>
- Choi SK, Adams SA, Eberth JM, et al. Medicaid coverage expansion and implications for cancer disparities. *Am J Public Health*. 2015;105(suppl 5):S706-S712. doi:10.2105/AJPH.2015.302876
- Corallo B, Moreno S. Analysis of recent national trends in Medicaid and CHIP enrollment. Kaiser Family Foundation. February 2, 2022. Accessed February 25, 2022. <https://www.kff.org/coronavirus-covid-19/issue-brief/analysis-of-recent-national-trends-in-medicaid-and-chip-enrollment/>
- Racial and ethnic disparities in Medicaid: an annotated bibliography. Medicaid and CHIP Payment and Access Commission. April 2021. Accessed February 25, 2022. <https://www.macpac.gov/wp-content/uploads/2021/04/Racial-and-Ethnic-Disparities-in-Medicaid-An-Annotated-Bibliography.pdf>
- Koroukian SM, Bakaki PM, Raghavan D. Survival disparities by Medicaid status. *Cancer*. 2012;118(17):4271-4279. doi:10.1002/cncr.27380
- Hinton E, Rudowitz R, Stolyar L, Singer N. 10 things to know about Medicaid managed care. Kaiser Family Foundation. October 29, 2020. Accessed January 25, 2022. <https://www.kff.org/web/2022021193245/>
- Share of Medicaid population covered under different delivery systems. Kaiser Family Foundation. Accessed December 10, 2021. <https://bit.ly/44sew7Y>
- Pennsylvania Cancer Registry. Pennsylvania Department of Health. Accessed December 6, 2021. <https://www.health.pa.gov/topics/Reporting-Registries/Cancer-Registry/Pages/Cancer%20Registry.aspx>
- PHC4 - 16 years of results. Pennsylvania Health Care Cost Containment Council. November 2002. Accessed November 21, 2022. <https://www.phc4.org/wp-content/uploads/phc4fyi14.pdf>
- Data reporting information for new facilities. Pennsylvania Health Care Cost Containment Council. Accessed February 28, 2022. <https://www.phc4submit.org/ProviderInformation.aspx>
- RUCA data: code definitions, version 2.0. Rural Health Research Center. Accessed February 25, 2022. <https://depts.washington.edu/urruca/ruca-codes.php>
- About the Neighborhood Atlas. Center for Health Disparities Research. Accessed February 25, 2022. <https://www.neighborhoodatlas.medicine.wisc.edu/>
- Pennsylvania Cancer Registry reporting manual. Pennsylvania Department of Health. Updated 2021. Accessed December 6, 2021. <https://www.health.pa.gov/topics/Documents/Reporting-Registries/PCR%20Reporting%20Manual.pdf>
- Medicaid expansion to the new adult group. Medicaid and CHIP Payment and Access Commission. Accessed October 21, 2022. <https://www.macpac.gov/subtopic/medicaid-expansion/>
- Hill SC, Abdus S, Hudson JL, Selden TM. Adults in the income range for the Affordable Care Act's Medicaid expansion are healthier than pre-ACA enrollees. *Health Aff (Millwood)*. 2014;33(4):691-699. doi:10.1377/hlthaff.2013.0743
- Norris L. Medicaid eligibility and enrollment in Pennsylvania. *HealthInsurance.org*. Accessed October 21, 2022. <https://www.healthinsurance.org/medicaid/pennsylvania/>
- Age and cancer risk. National Cancer Institute. Updated March 5, 2021. Accessed October 21, 2022. <https://www.cancer.gov/about-cancer/causes-prevention/risk/age#:~:text=The%20incidence%20rates%20for%20cancer,groups%2060%20years%20and%20older>
- Schrag D, Vrnig BA, Warren JL. Linking tumor registry and Medicaid claims to evaluate cancer care delivery. *Health Care Financ Rev*. 2009;30(4):61-73.

38. 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):mmrr.002.04.a01. doi:10.5600/mmrr.002.04.a01
39. 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. doi:10.1111/1475-6773.13114
40. Perrailton 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(3):703-711. doi:10.1111/1475-6773.13901
41. Meguerditchian AN, Stewart A, Roistacher J, Watroba N, Cropp M, Edge SB. 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. doi:10.1002/jso.21528
42. Thompson CA, Jin A, Luft HS, et al. Population-based registry linkages to improve validity of electronic health record-based cancer research. *Cancer Epidemiol Biomarkers Prev*. 2020;29(4):796-806. doi:10.1158/1055-9965.EPI-19-0882
43. Lau EC, Mowat FS, Kelsch MA, et al. Use of electronic medical records (EMR) for oncology outcomes research: assessing the comparability of EMR information to patient registry and health claims data. *Clin Epidemiol*. 2011;3:259-272. doi:10.2147/CLEPS23690
44. Coding pitfalls. Presented at: 2014-2015 North American Association of Central Cancer Registries Webinar Series; 2015. Accessed August 7, 2023. <https://crgc-cancer.org/wp-content/uploads/2013/10/Coding-Pitfalls-2015-Slides.pdf>
45. Gran B. A second opinion: rethinking the public-private dichotomy for health insurance. *Int J Health Serv*. 2003;33(2):283-313. doi:10.2190/BV3W-0JAR-R61K-6KAU
46. Buchmueller TC, Allen ME, Wright W. Assessing the validity of insurance coverage data in hospital discharge records: California OSHPD data. *Health Serv Res*. 2003;38(5):1359-1372. doi:10.1111/1475-6773.00181
47. Chapter X: data dictionary. North American Association of Central Cancer Registries. Accessed January 15, 2022. <http://datadictionary.naacccr.org/default.aspx?c=10&Version=22>
48. Verrill C. Assessing the reliability and validity of primary payer information in central cancer registry data. North American Association of Central Cancer Registries. Accessed October 21, 2022. <https://www.naacccr.org/wp-content/uploads/2016/11/Assessing-the-Reliability-and-Validity-of-Primary-Payer-Information-in-Central-Cancer-Registry-Data.pdf>
49. Chattopadhyay A, Bindman AB. Accuracy of Medicaid payer coding in hospital patient discharge data: implications for Medicaid policy evaluation. *Med Care*. 2005;43(6):586-591. doi:10.1097/01.mlr.0000163654.27995.fa

Visit [ajmc.com/link/89425](https://ajmc.com/link/89425) to download PDF and eAppendix



**eAppendix A. Payer categories in Pennsylvania Cancer Registry and PHC4 records**

<b>PCR Primary Payer at Diagnosis*</b>	<b>Re-coded Category</b>
01 Not Insured	Uninsured
02 Not Insured, Self Pay	Uninsured
10 Insurance, NOS [not otherwise specified]	Other
20 Private Insurance: Managed Care, HMO, or PPO	Private
21 Private Insurance: Fee-for-Service	Private
31 Medicaid	Medicaid
35 Medicaid-Administered through a Managed Care plan	Medicaid
60 Medicare without supplement, Medicare, NOS	Medicare
61 Medicare with supplement	Medicare
62 Medicare-Administered through a Managed Care Plan	Medicare
63 Medicare with private supplement	Medicare
64 Medicare with Medicaid eligibility	Medicare
65 TRICARE	Other
66 Military	Other
67 Veterans Affairs	Other
68 Indian/Public Health Service	Other
99 Insurance Status Unknown	Unknown

\*These primary payer categories are the standard NAACCR categories used by all NAACCR certified state cancer registries (<http://datadictionary.naacr.org/default.aspx?c=10&Version=22>)

<b>PHC4 Payer type</b>	<b>Re-coded Category</b>
Uninsured, Self Pay or Charity/Indigent Care	Uninsured
Medicaid, Fee for Service	Medicaid
Medicaid, HMO	Medicaid
Medicaid, HMO	Medicaid
Medicare, PPO	Medicare
Medicare, POS	Medicare
Medicare Part A or B, Fee for Service	Medicare
Medicare, HMO	Medicare
Government, PPO	Other
Government, Fee for Service	Other
Government, HMO	Other
Government, Unknown / Not Listed	Other
Government, PPO	Other
Government, Fee for Service	Other
Government, HMO	Other
Government, Unknown / Not Listed	Other
Blue Cross, PPO	Private
Blue Cross, POS	Private
Blue Cross, Fee for Service	Private
Blue Cross, HMO	Private
Blue Cross, Unknown / Not Listed	Private
Commercial, PPO	Private
Commercial, POS	Private
Commercial, Fee for Service	Private

Commercial, HMO	Private
Commercial, Workers' Compensation	Private
Commercial, Automobile	Private
Commercial, Unknown / Not Listed	Private
Blue Cross, PPO	Private
Blue Cross, POS	Private
Blue Cross, Fee for Service	Private
Blue Cross, HMO	Private
Blue Cross, Unknown / Not Listed	Private
Commercial, PPO	Private
Commercial, POS	Private
Commercial, Fee for Service	Private
Commercial, HMO	Private
Commercial, Workers' Compensation	Private
Commercial, Automobile	Private
Commercial, Unknown / Not Listed	Private
Uninsured, Self Pay or Charity/Indigent Care	Uninsured
Unknown / Not Listed, Unknown / Not Listed	Unknown
Unknown / Not Listed, Unknown / Not Listed	Unknown

Additional re-classification procedure for PHC4 payer variables: Because PHC4 populates three different payer variables for each record (paytype1, paytype2, paytype3), we also apply the following hierarchy rule to generate a single insurance variable containing mutually exclusive re-coded categories: Medicare → Medicaid → Other → Private → Uninsured → Unknown. For instance, a record having “Medicare” as Paytype1 and “Private” as Paytype2 would be assigned Medicare.

**eAppendix B. Any misclassification of Medicaid by age groups and years before and after Medicaid expansion**

	Older than 50			Younger than 50			Pre-Medicaid Expansion (2010-2014)			Post-Medicaid Expansion (2015-2016)		
	Marginal Effects	95% CI	P	Marginal Effects	95% CI	P	Marginal Effects	95% CI	P	Marginal Effects	95% CI	P
<b>Medicaid managed care</b>	3.2	-1.4, 7.9	0.173	1.5	-4.0, 6.9	0.599	3.6	-1.3, 8.5	0.146	-0.6	-6.6, 5.4	0.841
<b>Sex (ref=female)</b>												
Male	-0.4	-2.4, 1.7	0.715	1.1	-1.9, 4.2	0.471	0.6	-1.8, 3.0	0.64	-0.7	-3.4, 2.0	0.606
<b>Age (ref=under 45)</b>												
45-49				0.2	-2.3, 2.6	0.896	-0.2	-2.9, 2.4	0.856	0.4	-3.7, 4.6	0.838
50-54		Ref					0.8	-1.6, 3.2	0.523	-2.8	-6.2, 0.5	0.101
55-59	0.2	-1.6, 2.1	0.795				-0.5	-2.5, 1.5	0.611	0.7	-2.6, 4.0	0.689
60-64	4.2***	2.3, 6.0	<0.001				3.9***	1.6, 6.2	0.001	3.8*	0.1, 7.5	0.047
<b>Race/ethnicity (ref=non-Hispanic White)</b>												
Hispanic	-4.9	-12.9, 3.2	0.235	-5.7	-13.3, 2.0	0.144	-4.9	-13.5, 3.7	0.268	-6.3	-13.1, 0.4	0.064
Non-Hispanic Black	-8.3**	-13.8, -2.8	0.003	-6.3*	-12.3, -0.2	0.041	-7.1*	-13.0, -1.2	0.019	-8.9**	-15.2, -2.6	0.005
Non-Hispanic Other	-5.0	-12.9, 2.8	0.211	-3.6	-11.3, 4.1	0.358	-6.2	-15.0, 2.7	0.172	-1.7	-9.4, 6.1	0.678
Unknown	-0.3	-9.9, 9.4	0.958	-5.7	-16.5, 5.2	0.305	-4.9	-15.1, 5.2	0.341	5.3	-5.0, 15.6	0.314
<b>Cancer type (ref=other)</b>												
Lung	2.2	-0.5, 4.9	0.109	-0.7	-5.3, 4.0	0.777	1.9	-0.8, 4.6	0.168	1.8	-2.1, 5.7	0.367
Breast	2.5	-1.6, 6.5	0.233	4.1*	0.1, 8.0	0.042	3.8*	0.1, 7.4	0.042	2.6	-2.0, 7.1	0.273
Prostate	4.8*	0.6, 9.0	0.024	4.2	-9.4, 17.8	0.543	4.0	-1.5, 9.4	0.152	6.2	-0.1, 12.5	0.055
Colorectal	1.3	-1.7, 4.3	0.402	6.2**	2.4, 10.1	0.001	3.3*	0.2, 6.5	0.037	1.8	-2.2, 5.8	0.373
<b>Cancer stage (ref=Localized)</b>												
In situ	-1.7	-5.7, 2.3	0.405	-1.0	-7.9, 5.9	0.78	-3.7	-8.5, 1.2	0.139	2.6	-3.5, 8.8	0.4
Regional	-1.3	-3.5, 1.0	0.276	-1.6	-4.0, 0.9	0.211	-1.0	-3.3, 1.3	0.401	-2.3	-5.2, 0.7	0.135
Distant	-1.2	-4.0, 1.5	0.387	-2.8	-6.0, 0.5	0.101	-1.2	-4.3, 1.9	0.452	-3.1*	-5.9, -0.3	0.033
Unknown	1.5	-3.3, 6.3	0.534	2.4	-5.6, 10.4	0.562	2.8	-3.3, 9.0	0.365	-0.7	-7.6, 6.3	0.853
<b>Region (ref=Urban)</b>												
Large town	-6.1**	-10.5, -1.7	0.007	-5.2	-10.8, 0.4	0.067	-6.6**	-11.5, -1.7	0.009	-4.1	-10.5, 2.2	0.203
Small town	-8.9*	-15.8, -2.1	0.01	-10.7**	-17.8, -3.6	0.003	-11.3***	-17.7, -4.9	0.001	-5.2	-15.5, 5.1	0.323
Rural	-3.6	-11.0, 3.8	0.34	-5.3	-13.1, 2.5	0.183	-6.6	-13.2, 0.1	0.053	0.2	-8.2, 8.7	0.956
<b>ADI (ref=Quartile 4)</b>												
Quartile 1	11.8***	7.5, 16.1	<0.001	12.0***	5.9, 18.0	<0.001	11.7***	6.9, 16.5	<0.001	11.8***	6.5, 17.1	<0.001
Quartile 2	10.2***	5.8, 14.7	<0.001	10.3***	5.3, 15.3	<0.001	9.7***	5.1, 14.3	<0.001	11.4***	6.5, 16.4	<0.001
Quartile 3	3.5*	0.6, 6.3	0.017	4.2*	0.5, 7.8	0.027	3.0*	0.3, 5.6	0.031	5.1*	1.2, 9.0	0.01

Year of diagnosis (ref=2016)											
2010	2.5	-3.7, 8.7	0.433	3.2	-4.6, 11.0	0.421	-2.5	-6.7, 1.8	0.252		
2011	3.4	-1.1, 7.9	0.14	3.3	-3.7, 10.3	0.359	-1.8	-5.1, 1.5	0.289		
2012	1.9	-2.7, 6.5	0.419	-2.4	-9.3, 4.6	0.501	-4.8*	-9.0, -0.6	0.024		
2013	6.2**	1.8, 10.7	0.006	3.1	-3.2, 9.5	0.335	-0.2	-3.7, 3.3	0.912		
2014	6.5**	1.9, 11.0	0.005	3.5	-3.0, 10.0	0.292	Ref				
2015	1.7	-1.8, 5.2	0.339	0.7	-4.7, 6.0	0.804			1.3	-2.1, 4.6	0.461

Note: The marginal effects from the logit models were multiplied by 100 to represent percentage point changes in misclassification probabilities. The standard errors are clustered at the PHC4 facility level.

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

eAppendix C. Misclassification of Medicaid as private insurance by age groups and years before and after Medicaid expansion

	Older than 50			Younger than 50			Pre-Medicaid Expansion			Post-Medicaid Expansion		
	Marginal Effects	95% CI	P	Marginal Effects	95% CI	P	Marginal Effects	95% CI	P	Marginal Effects	95% CI	P
<b>Medicaid managed care</b>	13.1***	9.8, 16.4	<0.001	11.7***	8.1, 15.3	<0.001	14.0***	10.6, 17.4	<0.001	8.7***	4.8, 12.6	<0.001
<b>Sex (ref=female)</b>												
Male	-1.5	-3.2, 0.1	0.073	-1.3	-3.6, 1.1	0.291	-0.7	-2.4, 1.0	0.402	-2.7*	-5.1, -0.3	0.026
<b>Age (ref=under 45)</b>												
45-49				-1.1	-3.2, 0.9	0.286	-1.7	-3.8, 0.3	0.094	-0.5	-4.5, 3.5	0.802
50-54							-1.1	-2.9, 0.7	0.234	-4.1**	-7.2, -1.0	0.01
55-59	-0.8	-2.5, 0.8	0.323				-3.2***	-4.9, -1.6	<0.001	-2.0	-5.1, 1.1	0.198
60-64	-1.4	-3.3, 0.5	0.147				-2.8**	-4.6, -1.0	0.002	-4.4*	-7.8, -1.0	0.012
<b>Race/ethnicity (ref=non-Hispanic White)</b>												
Hispanic	-7.6***	-11.1, -4.1	<0.001	-10.5***	-14.1, -7.0	<0.001	-8.8***	-11.7, -5.8	<0.001	-9.0***	-13.1, -5.0	<0.001
Non-Hispanic Black	-2.5	-5.7, 0.8	0.14	-2.9	-6.9, 1.1	0.156	-1.9	-5.4, 1.6	0.297	-4.5*	-8.9, -0.2	0.04
Non-Hispanic Other	1.7	-3.3, 6.8	0.499	-2.4	-7.4, 2.5	0.335	-0.6	-6.5, 5.4	0.851	1.2	-5.0, 7.4	0.709
Unknown	-2.9	-10.6, 4.8	0.457	-5.1	-12.6, 2.4	0.184	-3.9	-11.6, 3.9	0.331	-3.4	-12.2, 5.4	0.445
<b>Cancer type (ref=other)</b>												
Lung	1.0	-0.9, 3.0	0.308	-2.6	-6.4, 1.2	0.188	0.0236	-2.2, 2.2	0.983	1.3	-1.9, 4.5	0.433
Breast	1.1	-1.9, 4.0	0.484	2.9	-0.6, 6.3	0.105	3.0	-0.4, 6.3	0.084	0.2	-3.4, 3.7	0.932
Prostate	2.0	-1.4, 5.5	0.252	1.2	-10.3, 12.7	0.842	1.8	-2.3, 6.0	0.383	3.1	-2.7, 9.0	0.291
Colorectal	2.1	-0.1, 4.4	0.066	3.8*	0.6, 7.1	0.021	2.0	-0.4, 4.5	0.1	4.1*	0.5, 7.7	0.027
<b>Cancer stage (ref=Localized)</b>												
In situ	-2.6*	-5.2, -0.1	0.043	-1.3	-6.2, 3.6	0.601	-3.7*	-6.8, -0.5	0.022	1.3	-3.7, 6.2	0.619
Regional	-0.2	-1.8, 1.3	0.763	0.2	-1.8, 2.2	0.84	-0.2	-1.8, 1.5	0.853	0.3	-2.3, 2.9	0.82
Distant	-0.8	-2.8, 1.1	0.409	-0.3	-2.9, 2.3	0.814	-0.2	-2.4, 2.0	0.877	-1.8	-4.4, 0.8	0.184
Unknown	-3.9*	-7.0, -0.8	0.015	-3.8	-9.8, 2.2	0.21	-5.1**	-8.7, -1.4	0.006	-1.2	-6.7, 4.2	0.664
<b>Region (ref=Urban)</b>												
Large town	-0.9	-4.7, 2.9	0.632	-0.0413	-4.46, 4.38	0.985	0.0152	-3.9, 3.9	0.994	-1.7	-6.2, 2.9	0.478
Small town	-1.1	-7.0, 4.8	0.714	-2.0	-7.9, 4.0	0.512	-3.5	-9.1, 2.2	0.228	3.1	-6.2, 12.4	0.513
Rural	-1.6	-8.3, 5.1	0.643	1.0	-7.2, 9.1	0.816	-1.3	-8.8, 6.3	0.745	0.3	-6.7, 7.3	0.932
<b>ADI (ref=Quartile 4)</b>												
Quartile 1	9.3***	6.1, 12.5	<0.001	8.2***	4.3, 12.1	<0.001	7.6***	4.8, 10.5	<0.001	11.4***	6.9, 15.9	<0.001
Quartile 2	1.9	-0.7, 4.4	0.149	3.1	-0.1, 6.4	0.058	1.3	-1.2, 3.7	0.303	4.4*	0.9, 7.9	0.013

Quartile 3	1.6	-0.6, 3.8	0.163	2.2	-0.8, 5.2	0.148	1.4	-0.8, 3.7	0.199	2.5	-0.6, 5.6	0.118
<b>Year of diagnosis (ref=2016)</b>												
2010	0.9	-3.7, 5.4	0.707	-0.7	-6.1, 4.6	0.788	0.2	-2.5, 2.9	0.871			
2011	1.4	-2.5, 5.4	0.477	-0.1	-4.9, 4.7	0.967	0.8	-1.9, 3.5	0.548			
2012	-1.3	-4.9, 2.3	0.48	-4.0	-8.1, 0.1	0.054	-2.2	-5.1, 0.7	0.131			
2013	0.4	-3.1, 4.0	0.808	-1.9	-5.6, 1.9	0.327	-0.5	-3.0, 2.0	0.696			
2014	1.5	-2.0, 5.1	0.394	-2.0	-6.9, 2.9	0.417						
2015	0.3	-2.5, 3.0	0.844	-1.5	-5.2, 2.2	0.429				-0.5	-3.3, 2.4	0.741

Note: The marginal effects from the logit models were multiplied by 100 to represent percentage point changes in misclassification probabilities. The standard errors are clustered at the PHC4 facility level.

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

**eAppendix D. Differences in 5-year mortality rate between Medicaid and private insurance using payer information in Pennsylvania cancer registry vs. PHC4 records**

Cancer	Stage	5-Year Mortality Rate						Difference in Disparities Estimates ([A]-[B])
		Using the Pennsylvania Cancer Registry			Using the PHC4 records			
		Medicaid	Private	Difference (Medicaid-Private) [A]	Medicaid	Private	Difference (Medicaid-Private) [B]	
Breast	Distant	64.1	52.9	11.2	65.2	51.2	14	-2.8
	Localized	4.3	1.9	2.4	4.7	1.7	3	-0.6
	Regional	18.5	8.6	9.9	16.9	8.4	8.5	+1.4
Colorectal	Distant	75	63.9	11.1	76.5	63.5	13	-1.9
	Localized	13.2	4.4	8.8	12.9	3.5	9.4	-0.6
	Regional	24.7	13.9	10.8	26.4	13.1	13.3	-2.5
Lung	Distant	90.8	87.5	3.3	91.4	87.2	4.2	-0.9
	Localized	24.6	14.5	10.1	25.9	12.6	13.3	-3.2
	Regional	57.8	49.7	8.1	58.1	49.3	8.8	-0.7
Prostate	Distant	58.4	52.1	6.3	61.1	50.2	10.9	-4.6
	Localized	8.3	1.7	6.6	7.2	1.6	5.6	+1.0
	Regional	6.5	4.8	1.7	8.7	4.5	4.2	-2.5