

---

# Linking Tumor Registry and Medicaid Claims to Evaluate Cancer Care Delivery

Deborah Schrag, M.D., M.P.H., Beth A. Virnig, Ph.D., and Joan L. Warren, Ph.D.

---

*The utility of Medicaid claims for studying cancer care is not known. Our objective was to evaluate how well Medicaid claims capture diagnostic and treatment information recorded by the California Cancer Registry (CCR). We compared cancer treatment from Medicaid claims with CCR data, using 1988-2000 cases matched with 1997-1998 Medicaid enrollment data. Medicaid claims corroborated diagnoses for 73 percent of breast and 68 percent of colorectal cancers in CCR. Medicaid claims confirmed surgery for 67 percent of CCR's breast cancers. We found that Medicaid claims have moderate sensitivity for identifying cancer diagnoses and surgery. Linked registry-Medicaid data can identify indigent patients and the timing of Medicaid coverage.*

## INTRODUCTION

An enormous body of research demonstrates that disparities in health care based on socioeconomic status are pervasive. Identifying strategies to evaluate and ultimately improve health care delivery for the poor is a national priority (Institute of Medicine, 2001; Hewitt and Simone, 1999; Bach et al., 2004; Jha et al., 2005; Lurie, 2005; Satcher et al., 2005; Shavers and Brown, 2002; Trivedi et al., 2005). Medicaid programs provide primary access to

health care for nearly 25 million indigent American adults between the ages of 18 and 64 and record itemized claims for services provided to many program enrollees based on diagnostic and procedure codes (Centers for Medicare & Medicaid Services, 2009).

In the early 1990s, the National Cancer Institute initiated a linkage between SEER's cancer registry data and the Medicare claims reported to the CMS (Warren, Klabunde, and Schrag, 2002; National Cancer Institute, 2009). This linkage provides a mechanism to evaluate cancer care for elderly age 65 and over and has become an important resource for cancer care research. While cancer registries provide a detailed snapshot characterizing cancer diagnoses and initial surgical and radiation treatment, they do not track patients longitudinally other than for vital status. Linkage with Medicare claims provides a longitudinal data source for most Americans over age 65. Therefore, linkage of registry and Medicare data is a powerful strategy for evaluating cancer care for older Americans. Our purpose was to determine whether a similar linkage between tumor registries and Medicaid data could create a useful resource for characterizing cancer care for indigent persons younger than 65. To accomplish this, we obtained a data set linking the California Cancer Registry (CCR) with California Medicaid files. We chose California both because the feasibility of linkage was clearly established (Perkins et al., 2001), and because of the large size of its Medicaid

---

Deborah Schrag is with the Dana Farber Cancer Institute. Beth A. Virnig is with the University of Minnesota School of Public Health. Joan L. Warren is with the National Cancer Institute. The statements expressed in this article are those of the authors and do not necessarily reflect the views or policies of the Dana Farber Cancer Institute, University of Minnesota School of Public Health, National Cancer Institute, or the Centers for Medicare & Medicaid Services (CMS).

Program—6.5 million enrollees of 36.8 million state residents (17.6 percent) or 2.6 million enrollees of 24.3 million adult residents aged 21-64 (10.7 percent) (California Department of Health, 2009). The large geographic area of the State also minimizes the extent to which patients cross state boundaries to receive care.

There are several motivating reasons for linking Medicaid enrollment and claims data with tumor registry data. First, racial and ethnic disparities in cancer mortality are pervasive and the underlying causes remain poorly understood. The extent to which race/ethnicity versus socioeconomic status influence cancer treatment and outcomes is difficult to ascertain at the population level because tumor registry data do not capture individual-level information about socioeconomic status. Since Medicaid enrollment is a good proxy for socioeconomic position, linkage to registry data creates an individual-level variable that may help to disentangle the complex interrelationship between race/ethnicity and social class and their relative influence on cancer treatment and outcomes.

Second, Medicaid claims data may supplement information collected by tumor registries about cancer treatment. Although registries capture comprehensive information about incident cancers and the use of initial surgery and radiation, they do not identify chemotherapy use or track outcomes other than survival. Because Medicaid covers chemotherapy and prescription drug treatments for its enrollees, linkage of registry data with Medicaid claims has the potential to identify utilization of appropriate cancer treatments and thereby care quality among the indigent. Just as is the case for Medicare, the degree to which Medicaid claims are informative depends on the number of enrollees covered by health plans which submit itemized claims for

reimbursement. Managed care plans that rely on capitated systems for reimbursement are unlikely to report itemized claims and therefore, heavy managed care penetration in Medicaid potentially limits the utility of Medicaid claims files for tracking longitudinal cancer care.

We took advantage of a preexisting linkage between CCR data and enrollment and claims data from California's Medicaid Program to explore the potential yield of constructing a CCR-Medicaid linkage to facilitate cancer-related health services research. Specifically, we sought to identify the numbers of cancer patients identified by such a linkage, and the extent to which Medicaid claims data might corroborate the cancer diagnoses and initial cancer treatments recorded by CCR.

## **METHODS**

### **Data Sources**

#### **California Cancer Registry**

CCR is a statewide population-based cancer surveillance system that collects information about all incident cancers diagnosed in California (California Department of Health and Human Services, 2005). The registry collects cancer type, stage at diagnosis, and surgical and radiation treatment as the first course of therapy as well as patient demographics and insurance status. Long-term treatment and recurrence are not collected, but subsequent cancers and survival data are captured.

#### **State of California Medicaid Claims**

California's Medicaid Program maintains enrollment records that include Social Security number, date of birth and month-by-month Medicaid eligibility.

Medicaid also maintains files with all health care claims submitted to the program for reimbursement. These claims cover inpatient, outpatient, long-term care, and prescription drug use, and include information on primary and secondary diagnoses and procedure codes using the standard ICD-9 and CPT schema for medical billing (Matchware Technologies, Inc., 1996; Bradley et al., 2005). For this analysis, we obtained the California Medicaid files detailing enrollment histories for 1997 and 1998, and all claims for treatments billed in 1998. We used this data set because it had previously been linked by the CCR working in cooperation with California Medicaid. The linkage used AutoMatch (Bradley, Given, and Roberts, 2001) software for a probabilistic linkage of Medicaid enrollment records for persons with valid Social Security Numbers to CCR records for persons diagnosed between 1988 and 2000. Social security number, date of birth, sex, and ZIP Code were used to link the two resources in a 3-pass matching algorithm that identified 161,645 persons with 174,682 tumors. We obtained an anonymized data set from this link. An earlier analysis based on this linkage evaluated Medicaid enrollment status based on the health insurance status recorded by the CCR and found that the registry record for insurance status has poor sensitivity 48 percent but good specificity 98 percent when compared to the “gold standard” Medicaid enrollment files from California (Centers for Disease Control and Prevention, 1999-2000). Therefore, we relied on the enrollment data from Medicaid and not the CCR to determine Medicaid status.

### **Assembly of Analytic Cohorts**

All analyses were restricted to persons aged 18 to 64 because cancer is rare among

children and most adults over age 65 are insured by Medicare. We used 1998 CCR data, Medicaid eligibility files for 1997 and 1998 and Medicaid claims files for 1998. To estimate the potential numbers of patients specific to cancer site that could be identified from a Medicaid-registry linkage as well as the timing of enrollment in relationship to date of cancer diagnosis, we considered the 6,800 unique subjects with a primary invasive cancer diagnosis reported to CCR in 1998 who were enrolled in Medicaid for at least 1 month during that year. This approach to cohort definition enabled us to assess Medicaid enrollment patterns for a minimum of 1 year prior to the month of diagnosis. Figure 1 illustrates the steps in linkage and cohort assembly.

### **Duration and Timing of Medicaid Enrollment**

Using the Medicaid personal summary file (PSF), we determined whether Medicaid enrollees with cancer were long standing beneficiaries, or, whether Medicaid enrollment was precipitated by the cancer diagnosis based on the month of enrollment. The file records monthly eligibility and enrollment and indicates the type of Medicaid managed care plan in which a beneficiary was enrolled. After determining that enrollment in Medicaid lacked any seasonal variation, we determined the timing of first enrollment in relation to the month of diagnosis and the total duration of enrollment over the 24-month interval between January 1997 and December 1998 for each patient whose cancer was diagnosed in 1998 and reported to the CCR.

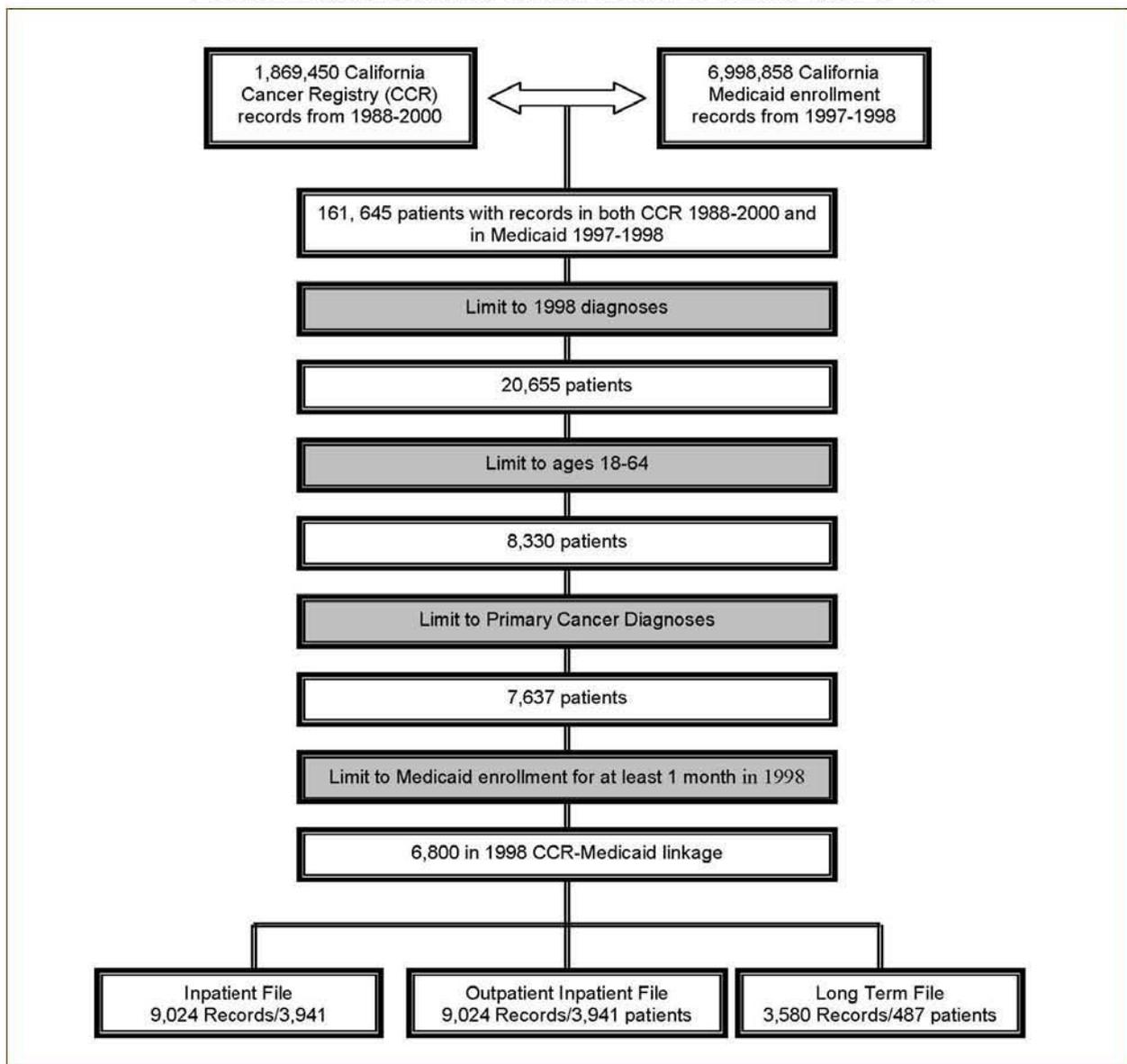
### **Type of Medicaid Coverage**

In California in 1997-1998, most Medicaid enrollees participated in some form of

managed care. However, in contrast to Medicare managed care plans which do not submit itemized claims to CMS, some Medicaid managed care plans do submit itemized claims (California Department of Health, 2009). The reason to distinguish between managed care plans and fee for service plans is that encounter data (captured in itemized claims indicating

specific health care services) permit assessment of longitudinal aspects of cancer care such as cancer chemotherapy. One approach would be to restrict analyses to the subgroup of Medicaid beneficiaries not covered by managed care plans. However, this would unnecessarily exclude many Medicaid enrollees who participate in managed care plans that do report en-

**Figure 1**  
**Flow Chart Describing Assembly of the CCR-Medicaid linked Cohort –**  
**Patients Diagnosed with a Primary Cancer in 1998 at ages 18-64.**



SOURCE: Schrag, D., Dana Farber Cancer Institute, Virnig, B.A., University of Minnesota School of Public Health, and Warren, J.L., National Cancer Institute, 2009.

counter data. Categorization of managed care status is further complicated by several features: (1) month-to-month variability in plan enrollment; (2) within plan variation such that some services may be covered and paid through a capitated rate while others are itemized; and, (3) lack of a variable in the Medicaid files that indicates if a plan reports encounter level data. Recognizing that no simple categorical definition of Medicaid “managed care” exists, we used the monthly “pre-paid health plan” variable recorded in the Medicaid PSF to identify enrollment. Because there is no easy way to distinguish among plans that do and do not submit claims data based on the information in the California Medicaid files, we searched for diagnostic and procedure codes for CCR-Medicaid enrollees irrespective of their participation in managed care. Then, we repeated our analyses restricting the analytic cohorts to Medicaid enrollees not enrolled in managed care (and therefore considered to be in a fee-for-service plan).

In California, as in other States, Medicaid enrollees younger than age 65 may obtain dual eligibility for Medicare if they qualify for disability benefits. For people who are eligible for full Medicaid coverage, the Medicaid Program supplements Medicare coverage by providing services and supplies that are available under their State’s Medicaid Program. Services that are covered by both programs will be paid first by Medicare. In cases where Medicare does not cover the full amount, the remainder (up to the State’s limit), is paid by Medicaid. We used identifiers in the Medicaid PSF indicating “Medicare crossover eligibility” (dual Medicare/Medicaid) and/or private insurance eligibility to identify Medicaid enrollees not expected to have informative claims files based on concurrent enrollment in Medicare. In this manner,

we eliminated patients with dual eligibility from our cohorts. We did not query Medicare claims directly. We included all other Medicaid enrollees aged 18-64 irrespective of whether the basis for eligibility was poverty, disability or special California entitlement programs.

## Medicaid Claims

In order to determine whether Medicaid claims submissions substantiate cancer diagnoses, we compared registry and claims derived diagnostic codes recorded in 1998. To consider the possibility that Medicaid claims corroborating CCR diagnoses appear with some time lag, we also examined records of the subgroup of cancer patients diagnosed in the first half of 1998 who were continuously enrolled in Medicaid for 6 months from diagnosis. In this manner, we were certain to have a minimum of 6 complete months of claims subsequent to diagnosis. We compared the cancer diagnosis code recorded by the CCR with the International Classification of Diseases-9 (ICD-9) (Centers for Disease Control and Prevention, 1999-2000) diagnosis codes recorded in the Medicaid inpatient, outpatient and long-term care files for the 6-month period starting with diagnosis.

In addition to assessing the diagnosis, we evaluated whether Medicaid claims submissions substantiated the information about initial cancer-directed surgery recorded by CCR. We restricted our analysis to breast and colorectal cancers, because initial treatment for these tumors almost always includes surgery which is reliably recorded by hospital-based CCR registries. We measured how frequently procedure codes recorded in Medicaid claims files during the 6-month window starting with the month of diagnosis corroborated information about primary

**Table 1**  
**Number of 1998 Incident Primary Cancers, By Site Reported to the California Cancer Registry (CCR) and the Proportion of Cancer Patients age 18-64 Enrolled in the California Medicaid Program**

Tumor Site	Incident Cancer Cases from the California Cancer Registry			Cases in both the California Cancer Registry and California Medicaid	
	All 1998 Cases	1998 Cases Age 18-24	% Cases Age 18-24	Cases Age 18-64 and in Medicaid for >=1 month in 1998	% of All 1998 Cases, Age 18-64 who had Medicaid for >=1 month in 1998
Breast	20,864	11,392	54.60%	1,014	8.90%
Lung	17,004	5,203	30.60%	997	19.20%
Colorectal	10,254	2,796	27.30%	557	19.90%
Cervical	1,690	1,350	79.90%	287	21.30%
Prostate	19,001	5,605	29.50%	268	4.80%
Testis	937	877	93.60%	78	8.90%
Uterine	3,587	1,722	48%	176	10.20%
Bladder	5,452	1,445	26.50%	113	7.80%
Hepatoma	1,570	836	53.30%	288	34.40%
Stomach	2,552	822	32.20%	197	24.00%
Ovarian	2,626	1,468	55.90%	188	19.60%
Kidney	2,914	1,358	46.60%	173	12.70%
Brain	1,922	1,246	64.80%	206	16.50%

SOURCE: Schrag, D., Dana Farber Cancer Institute, Virnig, B.A., University of Minnesota School of Public Health, and Warren, J.L., National Cancer Institute, 2009.

surgery recorded in CCR. For example, for breast cancer patients who had mastectomy and/or lumpectomy reported in the registry data, we queried the Medicaid claims for the corresponding ICD-9 and Current Procedural Terminology (CPT) procedure codes (ICD-9: 85.2, 85.3, 85.4 and CPT-11 19120, 19125-6, 19160-62, 19180-19240). We did not require exact matches between the procedure types recorded in CCR and Medicaid; rather, we categorized Medicaid claims as corroborating or not based on presence of any procedure code signifying a breast cancer operation. Biopsies were not considered as evidence of definitive breast surgery. Colon cancer operations were identified based on ICD-9 Codes: and CPT codes for total or partial colectomy (ICD-9 45.7, 45.8, 48.4, 48.5, 48.6 and CPT-11 codes 45397, 45110, 45126, 44140-44160).

In cases where analysis of Medicaid

claims did not identify the specific procedures corresponding to the cancer operation recorded in the CCR, we searched the inpatient Medicaid files to determine if there was any hospital admission recorded in 1998. If a hospital admission was recorded, we reviewed all procedure and diagnosis codes submitted for that admission to determine if any itemized claims were present and if present, whether they revealed evidence of cancer surgery.

### **Analytic Strategy and Approvals**

All analyses were descriptive and used SAS software version 9.0. Analyses of registry and claims data are Institutional Review Board (IRB) exempt, but signed Data Use Agreements stipulated standards for data protection and the study was reviewed by the CCR, California Medicaid,

**Table 2**  
**Duration and Timing of California Medicaid Enrollment in Relation to Cancer Diagnosis**

Number of Patients Recorded in CCR with a Primary Cancer Diagnosis in 1998 with Records in 1/97-12/98 Medicaid Files	Number of Patients (N=6,800)	Percent of Cohort
Medicaid Enrollment Status at month of diagnosis		
YES, Enrolled during month of diagnosis	5,351	79%
NO, First enrolled after month of diagnosis <sup>1</sup>	1,247	18%
NO, Enrolled prior to diagnosis but not in month of diagnosis	202	3%
Duration of Medicaid Enrollment prior to diagnosis		
>=12 months	3,323	49%
<12 but >=6 months	1,450	21%
<6 months	578	9%
Not enrolled until AFTER diagnosis	1,247	18%
Enrolled prior to diagnosis, but not during month of diagnosis	202	3%

<sup>1</sup>Of these 714 or 57%, enrolled in Medicaid within 3 months of diagnosis.

SOURCE: Schrag, D., Dana Farber Cancer Institute, Virnig, B.A., University of Minnesota School of Public Health, and Warren, J.L., National Cancer Institute, 2009.

and the Division of Cancer Control and Population Sciences at the National Cancer Institute. Analytic work conducted at Memorial Sloan Kettering Cancer Center (MSKCC), used encrypted SSNs and, a waiver of authorization was obtained from the MSKCC IRB.

## RESULTS

Table 1 illustrates the number of California residents, ages 18-64, who were diagnosed with select primary cancers in 1998 and the number of these who were enrolled in Medicaid for 1 or more months during the 1998 calendar year. Results provide site-specific estimates of the potential number of cases that would be identified by linking CCR and Medicaid. Notably, several tumors that are important causes of mortality for adults under age 65 are over-represented compared to their frequency in the population as a whole. For example, Medicaid beneficiaries constituted 34.4 percent of the hepatoma and 21.1 percent of the cervical cancer cases reported to CCR in 1998.

## Duration and Timing

A major threat to the utility of Medicaid data for health services research is the potential for discontinuous enrollment. However, as shown in Table 2, we found that the majority of patients identified in the linked data set were continuously enrolled in Medicaid prior to diagnosis. Among the 6,800 patients in the linked data set based on a 1998 cancer diagnosis at age 18-64, 79 percent were already enrolled in Medicaid during the month of diagnosis, 18 percent were enrolled in Medicaid only after the month of diagnosis and 3 percent were enrolled prior to diagnosis, but were no longer enrolled by the diagnosis month itself. When we restricted analyses to patients diagnosed in the first half of 1998 so as to ensure 6 months of Medicaid enrollment data subsequent to diagnosis, we found that 20 percent first enrolled in Medicaid after the diagnosis month. We found that among patients enrolled in Medicaid during the month of diagnosis, 3,323/5,351 (62 percent) had

**Table 3**  
**Cancer Diagnoses and Procedures in California Medicaid Claims for Patients with Cancer Diagnoses Recorded in CCR. Patients Were Enrolled in Medicaid for at Least Some Portion of 1998 and Age 18-64 at Diagnosis.**

Cancer Type	Diagnosed At Any Time During 1998		Diagnosed in the First Half of 1998	
	Number of CCR-Medicaid Enrollees with Cancer	Number (%) with a Corroborating Diagnosis Code Recorded in 1998 Medicaid Claims	Number of CCR-Medicaid Enrollees with Cancer	Number (%) with a Corroborating Diagnosis Code Recorded in 1998 Medicaid Claims Files
Breast	1,014	738 (73%)	549	409 (74%)
Lung	997	751 (75%)	523	416 (80%)
Colorectal	557	378 (68%)	274	189 (69%)
Cervical	287	207 (72%)	146	111 (76%)
Prostate	268	176 (66%)	127	93 (73%)
Testis	78	58 (77%)	37	28 (75%)
Uterine	176	121 (69%)	86	62 (72%)
Bladder	113	86 (76%)	53	42 (79%)
Liver	288	180 (63%)	151	95 (63%)
Stomach	197	137 (75%)	91	70 (77%)
Ovarian	188	130 (69%)	100	74 (76%)
Kidney	173	118 (68%)	93	70 (75%)
Brain	206	128 (65%)	98	64 (65%)

SOURCE: Schrag, D., Dana Farber Cancer Institute, Virnig, B.A., University of Minnesota School of Public Health, and Warren, J.L., National Cancer Institute, 2009.

been enrolled for at least 12 previous months. Once diagnosed with cancer, we found that it was rare for Medicaid beneficiaries to disenroll. For example, of the 3,116 patients diagnosed during the first half of 1998 who were also enrolled in Medicaid at diagnosis we found that 2938 (94 percent) remained continuously enrolled for the 6 months following diagnosis.

### Correspondence of Diagnostic Codes

As shown in Table 3, the diagnostic codes recorded on Medicaid claims corroborated the CCR diagnosis codes for 65-77 percent of patients, depending on the cancer site. For example, there were 1,014 primary breast cancer tumors reported by CCR for 18-64 year old Medicaid enrollees in 1998. For 738 (73 percent), a diagnosis code for breast cancer was found in 1998 claims.

To consider the possibility that diagnoses made late in calendar year 1998 were only reflected in 1999 claims files to which we did not have access, we examined claims for patients diagnosed during the first half of 1998 and found that the rate increased only marginally to 409/549 or 75 percent (Table 3).

In order to understand how often claims might represent incorrect diagnoses, we also looked at how often Medicaid indicated a cancer other than the one reported by CCR. For example, 1,014 of the 6,800 patients in our linked cohort had breast cancer reported by CCR. Only six patients of the remaining 5,786 (0.09 percent) had a breast cancer diagnosis represented on a claim but no CCR record of breast cancer. Across all cancer sites, the diagnosis codes in Medicaid claims only rarely suggested that a patient had a cancer from a site other than the

**Table 4**  
**The Sensitivity of California Medicaid Claims for Identifying Breast Cancer Surgeries among Women with Primary Breast Cancer Diagnoses and Breast Cancer Surgery in the First Half of 1998 Based on California Cancer Registry (CCR) Records**

Cohort Definition	Denominator Total N	YES, Corroborating Surgery Claim Identified		NO, Corroborating Surgery Claim Not Identified	
		N	%	N	%
How often was there a claim for breast surgery in the Medicaid records of women who had breast cancer diagnoses and breast surgery recorded in CCR?	467	331	71%	136	29%
Among the subgroup enrolled in a managed care plan?	240	156	65%	84	35%
Among the subgroup not enrolled in a managed care plan?	227	175	77%	52	23%
How often were there claims for breast surgery in 1998 Medicaid records of women continuously enrolled in Medicaid during 1998?	239	174	73%	65	27%
How often were there claims for breast surgery in 1998 Medicaid files for women enrolled for at least 1 but fewer than 12 months of 1998 in a Medicaid plan?	228	153	68%	75	32%
How often were there any inpatient Medicaid claims recorded for women who had primary breast cancer surgery recorded by the CCR, but no claim for breast surgery identifiable in Medicaid files?	136	45	33%	91	67%

SOURCE: Schrag, D., Dana Farber Cancer Institute, Virnig, B.A., University of Minnesota School of Public Health, and Warren, J.L., National Cancer Institute, 2009.

one reported by CCR; lung cancer was the most frequently reported cancer in Medicaid claims lacking substantiation in CCR (16/5,803=0.24 percent). Because we restricted our cohort to persons with primary cancer diagnoses reported to CCR, these rare instances where Medicaid claims identify a cancer not reported by the registry cannot be attributed to second tumors.

### Surgical Procedure Codes

We focused on patient cohorts with breast and colon cancer to evaluate the extent to which Medicaid claims for surgery corresponded to cancer-directed surgery recorded by CCR. We restricted this assessment to diagnoses made in the first half of 1998 to permit a minimum 6-month window from diagnosis for

breast cancer surgeries to be recorded in Medicaid files.

We identified 467 Medicaid-enrolled women who had breast cancer surgery recorded in CCR. We found that 331/467 or 71 percent, had a corroborating claim for breast surgery present in the 1998 Medicaid file. Among 240 of these 467 women enrolled in a managed care plan, 65 percent had Medicaid claims substantiating their breast surgery and 35 percent did not. Among the 227 not enrolled in managed care, 77 percent had claims substantiating their surgery and 23 percent did not (Table 4). Thus, managed care enrollment did not necessarily signify the lack of itemized claims and conversely, lack of managed care enrollment did not correspond to the consistent availability of claims. This result underscores the difficulty of using variables identifying

managed care status as a proxy for the presence or absence of encounter-level data in California Medicaid data. When these analyses were repeated for colon cancer, similar results were obtained; For example, claims did not reliably identify cancer surgery.

## DISCUSSION

We undertook an exploratory analysis to estimate the utility of linking California registry data with California Medicaid files for health services research and identified both opportunities and potential pitfalls. In California, the number of Medicaid-eligible cancer cases among adults aged 18-64 represents a meaningful proportion of the total incident cancers in this age bracket. This result is tempered by finding that California Medicaid claims are an imperfect source of data regarding cancer diagnoses and treatment. Overall, this highlights both the opportunities and challenges that researchers will encounter when trying to use these data for studying cancer care delivery for the poor.

It is important to understand limitations of existing systems with an eye toward their modification and improvement. Our analyses identify some of the reasons that Medicaid claims data currently fails to capture diagnoses and procedures. First, some patients with cancer under the age of 65 are simultaneously enrolled in Medicare and for these dually eligible, Medicare is typically the first payer. In such cases, Medicaid claims and for care that it covers where it pays 100 percent, Medicaid claims may not capture the use of specific health services because they were covered by and itemized on Medicare claims submissions. Another challenge is the lack of continuous enrollment, but even when analyses were restricted to the subset enrolled continuously, diagnosis

and procedure information was too often missing for this linkage to serve as a reliable resource to evaluate care quality.

Although we anticipated that enrollment in pre-paid health plans would mean that encounter-level data were not present, in fact, we found that many patients in pre-paid plans had itemized claims records and some in non-prepaid plans did not have itemized records. Review of these data and of Medicaid managed care programs in California suggests that plan-specific information is necessary to determine whether a particular plan itemizes claims. Moreover, the categorizations used to identify plan types in the PSFs as managed care or not, do not necessarily indicate whether encounter-level data are recorded. Thus, in order to work with merged Medicaid claims and registry data, investigators require plan-specific information about whether encounter-level data are submitted. This contrasts with Medicare claims files where health services researchers can restrict their cohorts to the approximately 80 percent of beneficiaries enrolled in fee-for-service plans for whom encounter-level data are available. Limiting Medicaid analyses to the fee-for-service beneficiaries is also disadvantageous because managed care penetration in most regions of the country is higher in Medicaid than in Medicare.

Previous analyses linking Medicaid enrollment data and tumor registry data have illustrated the potential of such linkages (Bradley, Gardiner, and Given, 2005; Bradley, Given, and Roberts, 2001, 2002, 2003). In particular, Bradley has established that linking Medicaid and registry files enables measurement of socioeconomic status as a determinant of cancer treatment and outcomes and has demonstrated the strong correlations between Medicaid enrollment post-cancer diagnosis, late-stage disease, and poor

rates of survival. Similar analyses have been previously performed by others (Perkins et al., 2001; Bradley, Koroukian et al., 2006, 2007; Siran et al., 2003; Chattopadhyay and Bindman, 2005). Koroukian found that the sensitivity of Medicaid claims for ascertaining breast cancer diagnoses reported to the Ohio Cancer Registry was only moderate (Siran et al., 2003; Chattopadhyay and Bindman, 2005). In addition, the positive predictive value was low secondary to a high rate of false positive diagnoses although addition of procedure codes for breast surgery helped identify true breast cancer cases. To date, the diagnostic and procedure codes on Medicaid claims have been utilized in cancer-related health services research to a limited extent, particularly when compared to the widespread use of Medicare claims for this purpose. Chattopadhyay and Bindman et al. have shown that California Medicaid records do not consistently corroborate hospitalizations reported to the State discharge registry system. These authors suggested that linkage of these sources could provide better monitoring of health care delivery for the poor. Both the potential for such systems and the need to minimize selection bias inherent in analyses of observational data collected as part of routine care delivery have been recognized (Hanratty et al., 2008; Terris, Litaker, and Koroukian, 2007).

Our analysis underscores that validation studies comparing treatment information gleaned from Medicaid claims to that recorded by tumor registries, hospital discharge abstracts, and medical records will be a necessary prerequisite to the use of Medicaid data as a resource to evaluate cancer care.

Our study has several important limitations: we evaluated data for select years with dated data and only for the State

of California; we had only one complete year of claims histories and two years of enrollment records; and, we did not have access to plan-specific requirements for submission of itemized claims, which might have made ascertainment of subgroups with uninformative claims more straightforward. Similarly, we did not have Medicare claims files to determine whether dual Medicare-Medicaid enrollees had claims evident in Medicare files.

Efforts to foster uniformity in submissions to CMS and Medicaid file structures, coupled with simplification of variables indicating whether or not itemized claims are present, would greatly facilitate the use of these data for health systems research. For example, linkage of cancer registry and Medicaid enrollment data would facilitate identification of the stage distribution at diagnosis of long-term Medicaid enrollees in comparison to the rest of the population. If Medicaid insurance coverage facilitates timely access to health care, its long-term enrollees should not have more advanced cancer stages at the time of diagnosis than persons insured by other means. If linkage to Medicaid claims could reliably identify utilization of colonoscopy and screening mammography for program enrollees or, appropriate use of adjuvant therapies, it could serve as an efficient and relatively low cost system for tracking care delivery. Given the limited extent to which Medicaid claims identified breast cancer surgery in our analysis, for the time being, claims must be regarded as an unreliable source for evaluating care quality. Despite the political, financial and logistical obstacles to restructuring Medicaid for the purpose of evaluating care, many factors argue for the need. The size of the Medicaid population and its budget, and the imperative to reduce health care disparities should motivate increased cooperation to develop an efficient and relatively economical

system to evaluate health care delivery for the poor.

We conclude that linking CCR data to Medicaid files can identify sizeable cohorts of cancer patients and their Medicaid enrollment status in relationship to diagnosis. This information has been and can continue to be used as an important factor in evaluating differences in disease severity, treatment, and survival between adults in the Medicaid Program and those who are insured through other means. We suggest that linkage of tumor registry data to Medicaid enrollment data should be routinely performed to augment information from the registry. The addition of these data can be used to monitor stage distribution, first course of treatment, overall and cause-specific survival among the indigent. However, until the sensitivity of Medicaid claims for identifying major components of care can be improved or better established for particular plan types, tumor registry data will remain the preferred source for identifying treatment information and it appears unlikely that Medicaid claims will characterize aspects of care not captured by tumor registries such as chemotherapy use. Given persistent gaps in cancer treatment outcomes based on socioeconomic status (Shavers and Brown, 2002; Trivedi et al., 2005), data systems should be designed such that merged Medicaid and registry data can be used to reliably track cancer care for the poor.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the California Cancer Registry, especially Mark Allen, for constructing the linkage and providing valuable advice and guidance.

## REFERENCES

- Bach, P.B., Pham, H.H., Schrag, D., et al.: Primary Care Physicians Who Treat Blacks and Whites. *New England Journal of Medicine* 351(6):575-584, August 5, 2004.
- Bradley, C.J., Given, C.W., and Roberts, C.: Late Stage Cancers in a Medicaid-Insured Population. *Medical Care* 41(6):722-728, June 2003.
- Bradley, C.J., Gardiner, J., Given, C.W., et al.: Cancer, Medicaid Enrollment, and Survival Disparities. *Cancer* 103(8):1712-1718, April 15, 2005.
- Bradley, C.J., Given, C.W., Roberts, C.: Disparities in Cancer Diagnosis and Survival. *Cancer* 91(1):178-188, January 1, 2001.
- Bradley, C.J., Given, C.W., and Roberts, C.: Race, Socioeconomic Status, and Breast Cancer Treatment and Survival. *Journal of the National Cancer Institute* 94(7):490-496, April 3, 2002.
- California Department of Health: California Medicaid. Internet address: <http://www.medi-cal.ca.gov> (Accessed 2009.)
- California Department of Health and Human Services: *The California Cancer Registry: Mission Statement*. 2005. Internet address: <http://www.ccrca.org/abouttheccr> (Accessed 2009.)
- Centers for Disease Control and Prevention: *International Classification of Diseases, Ninth Revision, Clinical Modification, Hospital & Payer (ICD-9-CM)*, 1999-2000.
- Centers for Medicare & Medicaid Services: Medicaid. Internet address: <http://www.cms.hhs.gov/home/medicaid.asp> (Accessed 2009.)
- Chan, J.K., Gomez, S.L., O'Malley, C.D., et al.: Validity of Cancer Registry Medicaid Status Against Enrollment Files. *Medical Care* 44(10):952-955, 2006.
- Chattopadhyay, A. and Bindman, A.B.: Accuracy of Medicaid Payer Coding in Hospital Patient Discharge Data: Implications for Medicaid Policy Evaluation. *Medical Care* 43 (6):586-591, June 2005.
- Hanratty, R., Estacio, R.O., Dickinson, L.M., et al.: Testing Electronic Algorithms to Create Disease Registries in a Safety Net System. *Journal of Health Care for the Poor and Underserved* 19(2):452-465, May 2008.
- Hewitt, M. and Simone, J.: *Ensuring Quality Cancer Care*. National Academies Press. Washington, DC. 1999.

Institute of Medicine: *Crossing the Quality Chasm: A New Health System for the 21st Century*. National Academies Press. Washington, DC. 2001.

Jha, A.K., Fisher, E.S., Li, Z., et al.: Racial Trends in the Use of Major Procedures Among the Elderly. *New England Journal of Medicine* 353(7):683-691, August 18, 2005.

Koroukian, S.M., Beaird, H., Madigan, E, et al.: End-of-Life Expenditures by Ohio Medicaid Beneficiaries Dying of Cancer. *Health Care Financing Review* 28(2):65-80, Winter 2006.

Koroukian, S.M., Xu, F., Beaird, H., et al.: Complexity of Care Needs and Unstaged Cancer in Elders: A Population-Based Study. *Cancer Detection and Prevention* 31(3):199-206, 2007.

Lurie, N.: Health Disparities—Less Talk, More Action. *New England Journal of Medicine* 353(7):727-729, August 18, 2005.

National Cancer Institute: *SEER-Medicare*. Internet address: <http://healthservices.cancer.gov/seer-medicare> (Accessed 2009.)

Perkins, C.I., Wright, W.E., Allen, M., et al.: Breast Cancer Stage at Diagnosis in Relation to Duration of Medicaid Enrollment. *Medical Care* 39 (11):1224-1233, 2001.

Satcher, D., Fryer, G.E., Jr., McCann, J., et al.: What If We Were Equal? A Comparison of the Black-White Mortality Gap in 1960 and 2000. *Health Affairs* 24(2):459-464, March-April 2005.

Shavers, V.L. and Brown, M.L.: Racial and Ethnic Disparities in the Receipt of Cancer Treatment. *Journal of the National Cancer Institute* 94(5):334-357, March 6, 2002.

Siran, M., Koroukian, G., Cooper, S., et al.: Ability of Medicaid Claims Data to Identify Incident Cases of Breast Cancer in the Ohio Medicaid Population. *Health Services Research* 38(3): 947-960, 2003.

Terris, D.D., Litaker, D.G., and Koroukian, S.M.: Health State Information Derived from Secondary Databases is Affected by Multiple Sources of Bias. *Journal of Clinical Epidemiology* 60(7):734-741, July 2007.

Trivedi, A.N., Zaslavsky, A.M., Schneider, E.C., et al.: Trends in the Quality of Care and Racial Disparities in Medicare Managed Care. *New England Journal of Medicine* 353(7):692-700, August 18, 2005.

Warren, J., Klabunde, C., and Schrag, D.: Overview of the SEER-Medicare Data: Content, Research Applications, and Generalizability to the United States Elderly Population. *Medical Care* 40:IV3-IV18, 2002.

---

Reprint Requests: Deborah Schrag M.D., M.P.H., Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02116. E-mail: [deb\\_schrag@dfci.harvard.edu](mailto:deb_schrag@dfci.harvard.edu)