Assessing the Validity of Insurance Coverage Data in Hospital Discharge Records: California OSHPD Data

Thomas C. Buchmueller, Mark E. Allen, and William Wright

Objective. To assess the accuracy of data on “expected source of payment” in the patient discharge database compiled by the California Office of Statewide Health Planning and Development (OSHPD).

Data Sources. The OSHPD discharge data for the years 1993 to 1996 linked with administrative data from the University of California (UC) health benefits program for the same years. The linked dataset contains records for all stays in California hospitals by UC employees, retirees, and spouses.

Study Design. The accuracy of the OSHPD data is assessed using cross-tabulations of insurance type as coded in the two data sources. The UC administrative data is assumed to be accurate, implying that differences between the two sources represent measurement error in the OSHPD data. We cross-tabulate insurance categories and analyze the concordance of dichotomous measures of health maintenance organization (HMO) enrollment derived from the two sources.

Principal Findings. There are significant coding errors in the OSHPD data on expected source of payment. A nontrivial percentage of patients with preferred provider organization (PPO) coverage are erroneously coded as being in HMOs, and vice versa. The prevalence of such errors increased after OSHPD introduced a new expected source of payment category for PPOs. Measurement problems are especially pronounced for older patients. Many patients over age 65 who are still covered by a commercial insurance plan are erroneously coded as having Medicare coverage. This, combined with the fact that during the period we analyzed, Medicare HMO enrollees and beneficiaries in the fee-for-service (FFS) program are combined in a single payment category, means that the OSHPD data provides essentially no information on insurance coverage for older patients.

Conclusions. Researchers should exercise caution in using the expected source of payment in the OSHPD data. While measures of HMO coverage are reasonably accurate, it is not possible in these data to clearly identify PPOs as a distinct insurance category. For patients over age 65, it is not possible at all to distinguish among alternative insurance arrangements.

Key Words. Health insurance, discharge data, measurement error

Health services researchers use hospital discharge data to investigate a wide range of research questions. Discharge data is the basis for nearly all empirical analyses of hospital competition (recent examples include Keeler, Melnick, and Zwanziger 1999; Zwanziger et al. 2000; Kessler and McClellan 2000; Capps et al. 2001; Town and Vistnes 2001) and hospital quality (see McClellan and Staiger 1999a, 1999b; Gowrisankaran and Town 1999) as well as a growing body of research on the effects of managed care on hospital behavior and patient outcomes (e.g., Chernew, Scanlon, and Hayward 1998; Spetz 1999). The insurance coverage of patients is a key variable for studies in these areas.

Perhaps the most commonly used source of discharge data is California's Office of Statewide Health Planning and Development (OSHPD). In the OSHPD dataset, information on insurance coverage comes from a variable describing "expected source of payment." Several recent papers provide examples of different ways this variable has been used. Chernew et al. (1998) use expected source of payment to identify health maintenance organization (HMO) enrollees to examine the effect of HMO coverage on the determinants of hospital choice. In a hospital-level analysis of the demand for nurses, Spetz (1999) uses the percentage of a hospital's patients with HMO coverage as a proxy for the effect of selective contracting. Town and Vistnes (2001) infer the attractiveness of different hospitals to patients from the results of multinomial logit regressions. Because a key assumption of their model is that patients used in the regressions have an unrestricted choice among hospitals, they use the data on expected source of payment to limit their sample to patients with fee-for-service (FFS) Medicare coverage. The results of the studies and others depend on the accuracy of the information on patient insurance coverage in the OSHPD file.

This paper assesses the validity of the OSHPD data on expected source of payment using a dataset that matches OSHPD's patient-level discharge with administrative enrollment data from the health benefits program of the

This research was supported by grant no. 030561 from the Robert Wood Johnson Foundation's Changes in Health Care Financing and Organizations program. We received valuable feedback from Candace Diamond, Diane Dargan, Starla Ledbetter, and Ginger Cox from the California Office of Statewide Health Planning and Development. All opinions and conclusions are those of the authors and should not be attributed to either the Foundation or OSHPD.

Address correspondence to Thomas C. Buchmueller, Graduate School of Management, University of California—Irvine, Irvine, CA, 92612-3125. Mark E. Allen is with the Public Health Institute, Berkeley, CA. William Wright is with the Cancer Surveillance Section, California Department of Health Services.

University of California (UC). The matched sample consists of UC employees, retirees, and spouses who were admitted to a California hospital between 1993 and 1996. Because over this period the UC program included a mix of HMO and non-HMO plans, the matched file allows us to test how accurately these categories can be distinguished in the OSHPD data. The importance of this assessment derives from the widespread use of these specific data. In addition, since the underlying issues related to data collection and verification are not unique to California, our analysis is also relevant to studies using discharge data from other sources.

DATA

The confidential hospital discharge files that we use contain each patient's Social Security number (SSN), date of birth, zip code of residence, and exact date of admission. This information, which is more detailed than that which is available in the public use discharge files, allows a precise match between the OSHPD and UC datasets. The UC source files consist of all individuals employed at the UC from 1993 to 1997, their spouses, and all UC retirees and their spouses as of 1997. The OSHPD data consists of all hospital discharges for the period from January 1, 1993, to December 31, 1996. For each year the OSHPD file contains between 3.6 and 3.7 million records. After removing records without SSNs there are between 2.7 and 2.8 million records per year. The UC files represent a total of 255,834 individuals, of whom 250,042 have SSNs.

We performed a probabilistic linkage between the two data sources using the *AutoMatch* software (Matchware Technologies 1997).¹ We accepted as a match all cases where the SSN, sex, and year of birth were an exact match, plus either the day or month of birth was also an exact match. In addition, in cases where there was an exact match according to the date of birth, sex, and first three digits of the zip code, we used the SSN and full zip code to estimate the likelihood that a pair of records was a match. Pairs of records with probability scores above a threshold were considered a match. For all of the matches made using this latter procedure there was either a single-digit difference or a transposition of two digits on the SSN. This matching procedure yielded a sample of 65,374 admissions for 39,032 individuals.

California hospitals are required by law to provide inpatient discharge data to OSHPD.² Information on the expected source of payment has been available in the patient discharge file since 1991. The procedures used for capturing and reporting data vary across California's approximately five

hundred hospitals according to each hospital's business processes and information technology systems. Data on payers may be captured by the admissions clerk, registration desk, billing, health information management, or other department. Data quality and accuracy may vary according to training (on content of the data as well as the information technology in use) provided to employees who capture payer information. In addition, technicalities of primary and secondary coverage available on a patient's insurance card or from the patient's understanding, may not be clear, complete, or consistent with the actual coverage and final source(s) of payment.

As shown in Table 1, from 1991 to 1994, the expected source of payment variable took 12 distinct values. The collection and coding of this information was revised in 1995 and again in 1999 (California Office of Statewide Health Planning and Development 2001). Between 1995 and 1998 there were 14 distinct payer categories. Given the great research interest in the effect of managed care on treatment and outcomes, a key question is how well the OSHPD data can be used to distinguish HMOs, PPOs (preferred provider organizations), and other forms of private insurance. Prior to 1995, the expected source of payment variable included a category for HMO coverage, but not a separate category for PPOs. A separate category for PPO coverage was introduced in 1995. In both periods, there are several categories for different public insurance programs. Since all of the individuals in our sample are known to be privately insured, we cannot assess how accurately the

Table 1: The Coding of "Expected Source of Payment" in the OSHPD Discharge File

<i>1991 to 1994</i>	<i>1995 to 1998</i>
1. Medicare	1. Medicare
2. Medi-Cal	2. Medi-Cal
3. Workers Compensation	3. Workers Compensation
4. Title V	4. County Indigent Programs
5. Other Governmental	5. CHAMPUS/CHAMPVA/VA
6. Blue Cross/Blue Shield	6. Other Governmental
7. Insurance Company	7. HMO
8. HMO/Prepaid Health Plan	8. PPO
9. Self-pay	9. Private Insurance
10. No Charge	10. Blue Cross/Blue Shield
11. Other Nongovernmental	11. Self-Pay
12. Medically Indigent Services	12. Charity Care
	13. No Charge
	14. Other Nongovernmental

OSHPD data captures the source of public coverage or whether or not a patient is uninsured. Thus, we focus on the distinctions made in the OSHPD data between different types of private insurance.

During the period of our study, UC employees were enrolled in four types of plans: (1) group HMO; (2) nongroup HMO; (3) PPO, and (4) point-of-service (POS). The group-model HMOs are the Northern and Southern California Kaiser plans. Five different nongroup HMOs (i.e., independent practice association [IPA] or network model plans) are represented in our data. We distinguish between group and nongroup HMOs because nearly all Kaiser members are treated at Kaiser hospitals, which serve only Kaiser patients, whereas nongroup HMOs contract with many hospitals, which serve all types of payers. Thus, we would expect the OSHPD data to be more accurate for Kaiser enrollees. That being said, since all of the HMOs in our data are organizations that were founded as HMOs (or prepaid group health plans), have long histories in California and offer little or no non-HMO coverage in the state, there should be little ambiguity as to how they should be coded.

Over the period of our study there were several other plans on the UC menu. Since the least managed option has different coinsurance rates for in- and out-of-network providers, it is technically a PPO, though its network is so large that it resembles a traditional indemnity plan. Two other PPOs, including a Blue Shield plan, are in the data for 1993 only. Another plan is in the data for the entire period, though its benefit design was not constant. Initially a traditional PPO (with a deductible and different coinsurance rates for in- and out-of-network providers), it was redesigned as a three-tier POS plan in 1996. Thus, the expected source of payment members of this plan should have been coded as "Insurance Company" in 1993 and 1994, and PPO when the new category was established in 1995. The OSHPD documentation does not provide definitive guidance as to how this plan should have been classified in 1996 when it was a POS plan.

The UC retirees choose from the same menu of plans as active employees. For those who are eligible for Medicare, the PPOs provide Medigap coverage. During the period of our analysis roughly half of UC retirees with Medicare coverage were enrolled in a risk product offered by one of the HMOs in the program. Since 1999, the OSHPD data definitions allow for Medicare beneficiaries with traditional and HMO coverage to be identified separately in the data. However, during the period covered by our analysis, hospitals should have reported Medicare as the expected source of payment for all Medicare beneficiaries, even those covered under managed care arrangements.

ANALYSIS

Non-Medicare Patients

We begin our analysis by cross-tabulating the expected source of payment as coded in the OSHPD file with the actual type of insurance coverage according to the UC data (Table 2). We focus first on individuals who, according to UC records, were not covered by Medicare. Note that this group includes some active workers who are over age 65 but are not eligible for Medicare because of their employment status. Separate tabulations are reported for the periods 1993–1994 and 1995–1996 because of the changes in the OSHPD coding scheme that occurred between 1994 and 1995.

For both time periods, the accuracy of the OSHPD data is greatest for patients with HMO coverage. In 1993–1994, over 80 percent of the observations for each type of HMO are coded correctly. For the group HMOs this figure is roughly the same in the later period, though for nongroup HMOs the accuracy declined with the introduction of the new category for PPOs.

The most common error for group HMO enrollees is to be coded as having Medicare coverage. Further inspection of the data reveals that almost all of these errors pertain to patients who are aged 65 or older. It appears that Kaiser hospitals have a policy of assigning Medicare as the expected source of

Table 2: OSHPD Expected Source of Pay by Actual Insurance Type: Non-Medicare Patients

<i>Plan Type (Actual)</i>	<i>OSHPD Expected Source of Payment</i>							
	<i>N</i>	<i>HMO</i>	<i>PPO</i>	<i>Private Insurance</i>	<i>BC/BS</i>	<i>Medicare</i>	<i>Other Public</i>	<i>No Insurance</i>
<i>1993–1994</i>								
Group HMO	3,948	83.5%	N/A	3.4%	0.4%	9.2%	2.2%	1.3%
Nongroup HMO	6,922	82.8	N/A	10.9	0.9	1.6	2.5	1.4
Blue Shield PPO	841	32.9	N/A	14.2	50.5	0.7	1.3	0.4
Other PPO	3,979	26.8	N/A	66.1	1.2	3.2	1.5	1.3
<i>1995–1996</i>								
Group HMO	3,534	82.0	2.6	1.3	0.2	8.8	3.0	2.2
Nongroup HMO	7,795	74.0	14.6	2.2	0.5	2.0	2.7	4.0
PPO	2,612	37.3	39.2	16.3	0.5	3.6	1.9	1.2
Point-of-Service	1,612	70.0	18.8	5.0	1.1	1.6	1.6	1.0

Notes: The sample includes active UC employees and their spouses and UC retirees and their spouses who are not covered by Medicare. The category labeled “Private Insurance” corresponds to the “Insurance Company” category in the OSHPD data prior to 1995 and the “Private Insurance” category thereafter. Figures sum to 100% across rows.

payment for all patients in this age group, regardless of whether they are covered through a Medicare or commercial contract. In our data, over 80 percent of Kaiser enrollees who are aged 65 or older but not covered by Medicare are coded in the OSHPD data as having Medicare. The same type of problem exists for the other HMOs, though it is less severe. Roughly one-third of nongroup HMO patients who are over age 65 but not in Medicare are coded as having Medicare.

The OSHPD data is less accurate for the other plan categories. For 1993–1994, we distinguish the Blue Shield PPO from the others because of the separate Blue Cross/Blue Shield category in the OSHPD data. Only half of Blue Shield enrollees are placed in that category. However, since for most analyses the distinction between Blue Cross/Blue Shield plans and other private plans is not important, it may be more relevant to examine the percentage of Blue Shield patients in either the Blue Cross/Blue Shield or “insurance company” categories.³ Using this broader categorization, 64.7 percent are coded accurately, which is very similar to the rate for the other PPO plans in 1993–1994 (66.1 percent).

Between one-quarter and one-third of PPO patients are mistakenly recorded as having HMO coverage in the earlier time period. After the introduction of the new PPO category, such errors were no less prevalent. In fact, in 1995–1996 PPO enrollees are only slightly more likely to be coded as being in a PPO than in an HMO (39 versus 37 percent); 16 percent are recorded as having some other type of private insurance. As a consequence of these errors and the misclassification of nongroup HMO members, only 40 percent of patients (1,025 out of 2,558 observations) coded in the OSHPD data as being in a PPO were actually in a PPO.

While the detailed breakdown by plan type is interesting, most studies on the effect of insurance type use broader categories—that is, HMOs versus other private plans. Thus, to the extent that errors in classification are within, rather than across, these broad categories they are of limited consequence. Therefore, we investigate the accuracy of a dichotomous HMO variable based on the information on expected source of payment.⁴ Specifically, we define the following variables:

$$H_s = 1 \text{ if a patient is coded as being in an HMO according to data source } s \\ = 0, \text{ otherwise, } \quad (s = U \text{ for UC, } O \text{ for OSHPD}).$$

We can then distinguish observations that are coded accurately in the OSHPD data ($H_U = H_O$) with false negatives ($H_U = 1$; $H_O = 0$) and false positives ($H_U = 0$; $H_O = 1$). Table 3 summarizes the results. Because we found the

expected source of payment variable to be especially error-ridden for individuals over age 65, we report separate results for patients above and below this age.

For 1993–1994, the sample HMO penetration rate according to the OSHPD variable is slightly lower than the actual rate (66.1 versus 69.3). Eighty percent of the observations are coded accurately, with false negatives slightly outnumbering false positives (11.7 to 8.6 percent). Because 93 percent of patients in our sample are under age 65, the percentages for the younger subsample are nearly the same as those for the full sample. In contrast, just over half of the older patients are coded accurately.

For 1995–1996, the results depend on how we classify the plan that went from being a PPO to a POS plan during this period. If we follow the standard convention of considering POS coverage as a type of HMO, 83 percent of our sample was in an HMO in 1995–1996. The OSHPD data does not capture this dramatic growth, as the HMO percent based on the expected source of payment variables increased only 3 percentage points, to 69.3 percent. If we leave the plan in the non-HMO category, the percent HMO according to the UC data is lower and hence so is the gap in the rates from the two sources. However, whichever approach we use, the rate of concordance is significantly lower in the later period as compared to the earlier one.

Sensitivity Tests

Our analysis is predicated on the assumption that insurance coverage is accurately recorded in the UC data, implying that any differences between the two data sources represent measurement error in the OSHPD data. One potential source of error in the UC data is the possibility that spouses of UC employees or retirees will be recorded as covered by a UC-sponsored plan, but will in fact use other insurance (i.e., through their own employer) when hospitalized. One way to test for such errors in the UC data is to compare results for employees and retirees selecting single coverage with individuals who are recorded as having two-party or family coverage. If patients who we believe to be covered through the UC are actually using other insurance when hospitalized, we should observe higher rates of concordance for single individuals since they are much less likely to have coverage from another source.⁵

Table 4 shows the results for the concordance of the measures of HMO coverage when we stratify the data in this way. For both time periods, there is no statistically significant difference in the rates of concordance between

Table 3: The Accuracy of HMO Coverage as Measured in the OSHPD Data

	1993-1994			1995-1996 (assuming POS= HMO)			1995-1996 (assuming POS ≠ HMO)		
	All	Age < 65	Age 65+	All	Age < 65	Age 65+	All	Age < 65	Age 65+
Percent of Patients in HMOs									
According to UC data (mean of H_U)	69.3%	70.0%	59.5%	83.2%	84.4%	68.7%	72.8%	74.0%	57.9%
According to OSHPD data (mean of H_O)	66.1	68.8	29.2	69.3	72.3	31.8	69.3	72.2	31.8
Percent coded accurately ($H_U=H_O$)	79.7	81.7	51.5	73.6	75.1	54.2	69.3	70.6	52.6
False positives ($H_U=0$; $H_O=1$)	8.6	8.5	9.1	6.3	6.4	4.5	13.6	13.8	10.7
False negatives ($H_U=1$; $H_O=0$)	11.7	9.8	39.4	20.2	18.5	41.3	17.1	15.6	36.7
Sample size	15,690	14,642	1,048	15,553	14,421	1,132	15,553	14,421	1,132

Notes: Samples consist of individuals who are not covered by Medicare. H_U = patient is coded as being in an HMO according to the UC data; H_O = patient is coded as being in an HMO according to the OSHPD expected source of payment variable.

Table 4: The Accuracy of the Coding of HMO Status by Coverage Tier

	<i>Employees and Retirees</i>	
	<i>With Single Coverage</i>	<i>With Coverage for Dependents</i>
<i>1993–1994</i>		
Number of observations	3,279	12,411
Percent correctly coded	79.2%	79.8%
<i>1995–1995</i>		
Number of observations	3,200	12,353
Percent correctly coded (POS = HMO)	68.7%	69.4%
Percent correctly coded (POS ≠ HMO)	73.4%	73.6%

individuals with single coverage and those with two-party or family coverage.

We also considered whether the results were sensitive to our coding of the UC plans. The results in Table 3 show that the percent HMO is sensitive to whether or not we treat the UC POS plan as an HMO, though the main qualitative results are largely unaffected. As noted, there is also some ambiguity about the coding of the least managed plan on the UC menu. We treat it here as a PPO, though because it has such a large network it could be viewed as a traditional plan, in which case “Private Insurance” would be the most accurate OSHPD category. Coding it in this way would clearly not affect the results of Table 2, since the coding of the HMO variable is not affected. Treating this plan as an indemnity plan would have a modest effect on the figures in Table 2 since by 1995 the plan enrolled fewer than 4 percent of all UC employees.

Medicare Beneficiaries

The final part of our analysis focuses on patients who, according to UC administrative records, were covered by Medicare. Recall that OSHPD coding guidelines during the period of our analysis called for Medicare to be listed as the expected source of payment for all beneficiaries, even those enrolled in an HMO. The figures in Table 5 indicate fairly close adherence to this rule. Over 90 percent of the observations for 1993–1994 were coded this way. For reasons that are not clear, the percentage fell slightly in 1995–1996. This change does not appear to be explained by the expansion of the OSHPD payment categories.

Table 5: OSHPD Expected Source of Pay by Actual Insurance Type: UC Retirees and Spouses with Medicare

<i>Plan Type</i>	<i>N</i>	<i>HMO</i>	<i>PPO</i>	<i>Private Insurance</i>	<i>BC/BS</i>	<i>Medicare</i>	<i>Other Public</i>	<i>No Insurance</i>
<i>1993–1994</i>								
Group HMO	2,274	3.8%	N/A	0.8%	0.0%	94.7%	0.6%	0.2%
Nongroup HMO	1,315	4.9	N/A	0.4	0.4	93.1	1.1	0.2
Blue Shield PPO	99	0.0	N/A	7.1	0.0	91.9	1.0	0.0
Other PPO	5,423	0.9	N/A	1.0	0.1	97.6	0.3	0.2
<i>1995–1996</i>								
Group HMO	2,238	6.8	0.6	0.7	0.0	91.6	0.2	0.1
Nongroup HMO	1,776	11.3	2.9	0.9	0.0	83.9	1.0	0.1
Point-of-Service	198	5.1	2.0	0.0	0.0	91.4	1.0	0.5
PPO	5,494	1.4	0.4	0.5	0.2	97.0	0.3	0.3

DISCUSSION

The typology used to describe health insurance plans is imprecise, with terms like HMO and “managed care” used to describe a broad range of payment mechanisms, contractual arrangements, and incentive structures. Among HMOs there exists significant diversity with respect to these dimensions, and differences between less tightly integrated HMOs and PPOs are often quite subtle. Given the inconsistent way these labels are often used, our finding that HMO coverage is measured with significant error in the OSHPD data should not come as a great surprise.

A number of analyses based on the OSHPD data have used the expected source of payment variable to identify patients with HMO coverage. For the years 1993 and 1994, a dichotomous HMO variable is incorrect for one-fifth of our matched sample of patients with private insurance coverage. The accuracy of the variable declined after the introduction of a separate payment category for PPOs in 1995 as a large number of HMO patients were incorrectly classified as being in PPOs. The noise created by these errors means that analyses that compare HMO and non-HMO patients will be subject to bias due to the contamination of the two comparison groups. We would not conclude from our results that such comparisons are not valid, but rather that researchers should recognize the potential bias resulting from this measurement error.

Since there are fundamental differences between HMOs and PPOs, it would appear that the finer insurance categories introduced by OSHPD in 1995 would allow for a more precise definition of insurance status and, therefore, richer analyses of the effects of organizational form on patient

outcomes. However, the results from our matched sample suggest that a PPO identifier based on the expected source of payment variable is essentially meaningless, at least in the first year that the new category was available.⁶ Thus, we would recommend against attempting to identify PPO patients as a separate group in the OSHPD data. The HMO penetration rates calculated at the hospital level may still be a reasonable proxy for the impact of selective contracting, though intertemporal changes in this variable for periods spanning the change in the expected source of payment variable are suspect.

We conclude that OSHPD discharge data from the 1990s provide little or no useful information on insurance coverage for patients over age 65. Part of this is a direct result of OSHPD's coding guidelines, which during the period of our analysis called for all Medicare beneficiaries in the FFS program and those enrolled in HMOs to be grouped in a single expected source of payment category.⁷ In addition, we find that a significant fraction of patients who are over age 65 but who are covered by a commercial plan are incorrectly coded as having Medicare coverage. As a result, we recommend that researchers using the expected source of payment variable should exclude all individuals over age 65, not just those identified as being covered by Medicare.

The lack of information on HMO coverage for Medicare beneficiaries is unfortunate given the high percentage of California seniors enrolled in capitated plans and the importance of research questions relating to the effect of that coverage. Revisions to the variable definitions enacted in 1999 should improve this situation considerably. Under current guidelines, expected source of payment is captured by two variables, the first of which pertains to payer category (nine categories including private, Medicaid, and Medicare) and the second, which pertains to payer type (four categories including HMO, other managed care, and traditional). This modification represents a potentially important improvement in the data. Assuming that the coding is accurate, it should now be possible to use the OSHPD discharge data to contrast outcomes for Medicare beneficiaries in HMOs with those in the fee-for-service sector. Of course, given the results of this study the accuracy should not be taken for granted.

The main limitations of our analysis come from the fact that our data are not representative of all California hospital patients. The most obvious implication is that we cannot examine how accurately Medicaid and uninsured patients are identified using OSHPD's expected source of payment variable. An analysis using a matched sample with patients from all payer types is an important subject for future work. Moreover, our sample is not representative of all privately insured patients. We would speculate that differences between our sample and the statewide population are such that our

results overstate the accuracy of the OSHPD expected source of payment variable during the period of our analysis. One reason is that compared to all privately insured Californians, a greater percentage of UC employees was in HMOs, and a smaller percentage was in PPOs, and expected source of payment was coded more accurately for HMO enrollees. In addition, the HMOs in the UC program are plans that have long histories in California and do very little non-HMO business. Coding errors are likely to be more common for multiproduct insurers that are less well known in the state.

At the same time, it is likely that the quality of the payment variable has improved since the time period covered by our data as hospital staff became more familiar with the OSHPD guidelines and with the changing structure of the health insurance market. The subsequent revisions of the definitions, in particular those relating to Medicare coverage, should also contribute to an improvement in data quality.

NOTES

1. The match was done at the California Cancer Registry. The enrollment data file with employee SSNs was returned to the UC after the project was complete. The project was approved by institutional review boards at the Cancer Registry and the University of California, Irvine.
2. The reporting of discharge data to OSHPD is required by the California Health and Safety Code and further defined in Section 97232 of the California Code of Regulations. Extensive details about the data are available from <http://www.oshpd.state.ca.us>.
3. For example, Chernew et al. (1998) combine the patients in the Blue Cross/Blue Shield categories with those in the "insurance company" category and compare them to patients for whom the expected source of payment is HMO/Prepaid Health Plan. In contrast, Capps et al. (2001) include Blue Cross/Blue Shield as a separate payment category in their regressions.
4. In the analysis by Spetz (1999), the county-level mean of this variable is used to measure market-level HMO penetration. Capps et al. (2001) report the mean for a patient-level variable labeled "HMO/PPO." However, since they use OSHPD data from 1991, we assume that this variable is the same as the HMO indicator variable we use here.
5. It is important to note that we only observe coverage choices, not actual marital status or family composition. Thus, while all single individuals should be observed with single coverage, it is possible for a married person to opt for single coverage if her spouse is covered elsewhere. Such cases could only lead to errors in the UC data if the UC employee chooses to renounce her UC coverage in favor of coverage through her spouse. However, coordination of coverage rules used by insurers will generally not allow this. Thus, there is little reason to suspect that individuals with single coverage through the UC actually use other insurance when hospitalized.

6. Because of the change in benefit design for the largest UC PPO, it is not possible with our data to track the accuracy of the OSHPD information on PPO coverage over a longer period of time.
7. Two recent studies (Town and Vistnes 2001; Capps et al. 2001) that use OSHPD data to model patients' hospital choice decisions limit their analysis to Medicare patients (and exclude HMO patients) on the assumption that Medicare beneficiaries face a constant price across all hospitals and their choices are not constrained by HMO selective contracting. That is, they assume that patients for whom Medicare is listed as the expected source of payment are in the FFS program. This assumption is clearly unjustified in light of OSHPD coding guidelines and the high percentage of California Medicare beneficiaries enrolled in HMOs.

REFERENCES

- California Office of Statewide Health Planning and Development (OSHPD). 2001. *2000 Patient Discharge Data File Documentation*.
- Capps, C. S., D. Dranove, S. Greenstein, and M. Satterthwaite. 2001. "The Silent Majority Fallacy of the Elzinga-Hogarty Criteria: A Critique and New Approach to Analyzing Hospital Mergers." Working paper no. 8216 (April), National Bureau of Economic Research, Cambridge, Mass.
- Chernew, M., D. Scanlon, and R. Hayward. 1998. "Insurance Type and Choice of Hospital for Coronary Artery Bypass Graft Surgery." *Health Services Research* 33 (1, part 1): 447-66.
- Gowrisankaran, G., and R. J. Town. 1999. "Estimating the Quality of Care in Hospitals Using Instrumental Variables." *Journal of Health Economics* 18 (6): 747-67.
- Keeler, E., G. Melnick, and J. Zwanziger. 1999. "The Changing Effects of Competition on Non-Profit and For-Profit Hospital Pricing Behavior." *Journal of Health Economics* 18 (1): 69-86.
- Kessler, D., and M. McClellan. 2000. "Is Hospital Competition Socially Wasteful?" *Quarterly Journal of Economics* 115 (2): 577-615.
- McClellan, M., and D. Staiger. 1999a. "Comparing Hospital Quality at For-Profit and Not-for-Profit Hospitals." Working paper no. 7324 (August), National Bureau of Economic Research, Cambridge, Mass.
- . 1999b. "The Quality of Health Care Providers." Working paper no. 7327 (August), National Bureau of Economic Research, Cambridge, Mass.
- Matchware Technologies. 1997. *AutoMatch* (release 4.1). Silver Spring, MD: Matchware Technologies.
- Spetz, J. 1999. "The Effect of Managed Care and Prospective Payment on the Demand for Hospital Nurses: Evidence from California." *Health Services Research* 34 (5, part 1): 993-1010.
- Town, R. J., and G. Vistnes. 2001. "Hospital Competition in HMO Networks." *Journal of Health Economics* 20 (5): 733-53.
- Zwanziger, J., G. A. Melnick, and A. Bamezai. 2000. "The Effect of Selective Contracting on Hospital Costs and Revenues." *Health Services Research* 35 (4): 849-67.