

# UCLA

## UCLA Previously Published Works

### Title

Improving Hospital Reporting of Patient Race and Ethnicity - Approaches to Data Auditing

### Permalink

<https://escholarship.org/uc/item/4c37x5cf>

### Journal

Health Services Research, 50(S1)

### ISSN

0017-9124

### Authors

Zingmond, DS  
Parikh, P  
Louie, R  
et al.

### Publication Date

2015-08-01

### DOI

10.1111/1475-6773.12324

Peer reviewed

© Health Research and Educational Trust  
DOI: 10.1111/1475-6773.12324  
ENHANCED HOSPITAL DISCHARGE DATA

# Improving Hospital Reporting of Patient Race and Ethnicity—Approaches to Data Auditing

*David S. Zingmond, Punam Parikh, Rachel Louie, Daphne Y. Lichtensztajn, Ninez Ponce, Romana Hasnain-Wynia, and Scarlett Lin Gomez*

---

**Objective.** To investigate new metrics to improve the reporting of patient race and ethnicity (R/E) by hospitals.

**Data Sources.** California Patient Discharge Database (PDD) and birth registry, 2008–2009, Healthcare and Cost Utilization Project’s State Inpatient Database, 2008–2011, cancer registry 2000–2008, and 2010 US Census Summary File 2.

**Study Design.** We examined agreement between hospital reported R/E versus self-report among mothers delivering babies and a cancer cohort in California. Metrics were created to measure root mean squared differences (RMSD) by hospital between reported R/E distribution and R/E estimates using R/E distribution within each patient’s zip code of residence. RMSD comparisons were made to corresponding “gold standard” facility-level measures within the maternal cohort for California and six comparison states.

**Data Collection.** Maternal birth hospitalization (linked to the state birth registry) and cancer cohort records linked to preceding and subsequent hospitalizations. Hospital discharges were linked to the corresponding Census zip code tabulation area using patient zip code.

**Principal Findings.** Overall agreement between the PDD and the gold standard for the maternal cohort was 86 percent for the combined R/E measure and 71 percent for race alone. The RMSD measure is modestly correlated with the summary level gold standard measure for R/E ( $r = 0.44$ ). The RMSD metric revealed general improvement in data agreement and completeness across states. “Other” and “unknown” categories were inconsistently applied within inpatient databases.

**Conclusions.** Comparison between reported R/E and R/E estimates using zip code level data may be a reasonable first approach to evaluate and track hospital R/E reporting. Further work should focus on using more granular geocoded data for estimates and tracking data to improve hospital collection of R/E data.

**Key Words.** Data auditing, race/ethnicity, gold standard comparisons

---

Racial and ethnic health disparities are well documented (Institute of Medicine 2003; Kim et al. 2011, 2012; Agency for Healthcare Research and Quality 2012). The Institute of Medicine recommends that to improve quality of care across racial/ethnic groups that valid and reliable data on race/ethnicity must be collected (Institute of Medicine 2009). Quality improvement efforts to reduce disparities in care across groups rely upon the existence of valid and reliable measurements of patient characteristics, including race and ethnicity (R/E). The National Quality Forum now recommends that future performance measures be stratified—or calculated separately—by sociodemographic factors, including income, race, and education (National Quality Forum 2014). The Affordable Care Act requires standardized collection of race/ethnicity across federal health care databases precisely for this reason (The Patient Protection and Affordable Care Act 2010). Hospital medical records data on patient sociodemographic characteristics serve as the foundation for identifying disparities in care and disease outcomes within and across medical systems. They are also the primary source of patient information for disease-based databases, such as cancer registries, which are the basis for identifying disparities in cancer occurrence and survival (Glaser et al. 2005).

Despite their importance, hospital medical record data have proven to be problematic sources of demographic data. Several studies have shown that medical record data on R/E are subject to misclassification (Stewart et al. 1999; Kressin et al. 2003; Gomez and Glaser 2005; Gomez et al. 2005). Questions remain as to the consistency in collection of these data within and across hospitals (Stewart et al. 1999; Kressin et al. 2003; Gomez and Glaser 2005; Gomez et al. 2005). Efforts have been made to improve self-reported data, including periodic contact via postcard to elicit R/E (Arday et al. 2000) and introduction of the National Consumer Assessment of Health Plans (Morales et al. 2001). However, because the collection of valid and reliable self-reported data on R/E in health care continues to lag, analytic approaches to improve the accuracy of these measures have been attempted, including

---

Address correspondence to David S. Zingmond, M.D., Ph.D., UCLA School of Medicine, Division of General Internal Medicine and Health Services Research, 911 Broxton Avenue, Los Angeles, CA 90024; e-mail: dzingmond@mednet.ucla.edu. Punam Parikh, M.P.H., and Rachel Louie, M.S., are with the UCLA School of Medicine, Division of General Internal Medicine and Health Services Research, Los Angeles, CA. Daphne Y. Lichtensztajn, M.D., and Scarlett Lin Gomez, Ph.D., M.P.H., are with the Cancer Prevention Institute of California, Fremont, CA. Ninez Ponce, M.P.P., Ph.D., is with the UCLA Fielding School of Public Health, Department of Health Services, Los Angeles, CA. Romana Hasnain-Wynia, M.S., Ph.D., is with the Patient-Centered Outcomes Research Institute, Washington, DC.

name-matching techniques to identify Hispanic and Asian-Pacific Islander enrollees (Morgan, Wei, and Virnig 2004; Wei et al. 2006; Eicheldinger and Bonito 2008) and Bayesian techniques (Elliott et al. 2008, 2009).

Although great effort has been expended to develop processes to improve the validity and reliability of self-reported R/E in the health care setting, a major weakness has been a lack of metrics to track and feedback the accuracy and completeness of R/E data collected by hospitals. For example, in California, the Office of Statewide Health Planning and Development (OSHPD) examines hospital discharge data for completeness (low rates of “other” or “unknown” reported) and for internal consistency (patients admitted to the same hospital have the same R/E across encounters). Unfortunately, these checks are not equivalent to accuracy, which requires comparison to gold standard (self-reported) R/E information. These self-report patient-level data are harder to obtain. Nationwide, few, if any, organizations that manage the respective statewide hospital data are currently making this a part of their data improvement efforts.

In order to improve data quality, we attempted to create a measure of overall accuracy of hospital reported R/E using the California inpatient data linked to the US Census, which could be used with all-payer hospital discharge datasets collected in California and other states. The hospital measure was validated through comparison to a gold standard–derived measure of accuracy. Finally, we attempted to use the metric to assess trends in data accuracy (new metric) and in data completeness (rate of missing/unknown) in California and six other demographically diverse states that submit data to the national Healthcare Cost and Utilization Project (HCUP). Our overarching goal was to create a validated measure that could be easily employed using existing data with the recognition that as more detailed information becomes available, more refined measures will create better estimates of hospital reporting.

## METHODS

### *Hospital Data*

To develop and implement new approaches to evaluating the accuracy of R/E and reporting in statewide hospital discharge data, we examined the California inpatient data. OSHPD is responsible for the routine collection of patient-level hospital inpatient, emergency department, and ambulatory surgery data in California. The inpatient file—OSHPD’s Patient Discharge Database

(PDD)—has over 4 million discharges annually and contributes to HCUP. To test the utility of measures developed, we used analogous data from six other states that contribute to HCUP (Washington [WA], Oregon [OR], Arizona [AZ], Colorado [CO], Florida [FL], and New Jersey [NJ]) for the years 2008–2011. These states were selected because they collect R/E, report data in a timely manner to HCUP, have reasonably sized diverse populations, and have differences in reporting requirements and data improvement efforts.

Race and ethnicity are required to be reported for each patient discharged from California hospitals. Race consists of six categories (white, Black, American Indian/Native Alaskan, Asian-Pacific Islander, other, and unknown). Ethnicity has three categories (Hispanic, not Hispanic, unknown). Multiple R/E responses are not allowed in the California data. Multiple race is directed to be classified as “other.” Ideally, unknown race or ethnicity should occur either if the respondent (the patient or patient proxy) does not know or if the patient cannot answer (e.g., comatose) and there is no proxy. Race and ethnicity are meant to be self-reported. Currently, there is no standard for confirming the source of reporting this information. Recent and previous surveys of hospital practices suggest that front line staff often assigned R/E by patient name and appearance (Hasnain-Wynia and Baker 2006).

### *Gold Standard*

Self-reported R/E are the gold standard for assessing the accuracy of reported R/E. We used two routinely collected sources of patient-level data with self-reported race/ethnicity as gold standards linked to the PDD for comparison—data from birth certificates and data from a cancer epidemiology database. A birth certificate is required for each live birth in California, is completed by the birth mother during the birth hospitalization, and includes the mother’s R/E. OSHPD has routinely linked maternal birth hospitalization records with the state birth registry maintained by the Office of Vital Statistics of the Department of Public Health (Office of Statewide Health Planning and Development 2009). In addition to linkage to the index hospitalization associated with a mother’s admission for labor and delivery, preceding and subsequent hospitalizations for the mother can also be examined.

Although the state cancer registry—the California Cancer Registry (CCR)—cannot be used for gold standard comparison to the PDD because its demographic data are abstracted mainly from the same hospital registration materials as the PDD, two of its regional registries—the Cancer Prevention Institute of California (CPIC; San Francisco Bay Area regional registry) and

the University of Southern California (USC; the Los Angeles County registry)—have created an epidemiologic database for the subset of cancer patients that have enrolled in cancer studies and whose self-reported demographic data have been reported back to CPIC and USC by cancer investigators. Although more selective and significantly smaller than the overall CCR, these data allow for examination of a different sample than the maternal data. Because of the smaller number of cancer cases, these data are best used for overall gold standard comparison, rather than within institution comparisons.

### *Proxy Gold Standard*

In contrast to routinely collected statewide patient-level data, Census 2010 Summary File (SF) 2 provides summary counts of actual self-reported and proxy-reported R/E by geographic unit. The Census allows for detailed R/E responses, including multiple R/E responses for individuals of mixed backgrounds. The Census Bureau follows the general Office of Management and Budget recommendations for multirace reporting and “rolling up” categories to the single R/E as well as providing counts for combined race/ethnicity.

We linked the Census 2010 SF2 summary race, ethnicity, and combined race/ethnicity count data to the PDD. For each hospital discharge record, we used the reported patient zip code to link to the corresponding Census zip code tabulation area (ZCTA). Patient zip codes without a corresponding physical location (e.g., post office boxes) were assigned to the closest zip code with a corresponding ZCTA. Zip codes with a corresponding physical location, but no corresponding ZCTA (usually in rural areas) were mapped to the closest ZCTA. Hospital discharge abstracts without a zip code were assigned one from among patients with zip codes discharged from that same hospital using hot deck imputation. If a hospital had no valid zip codes reported among its discharges, the hospital zip code (and corresponding ZCTA) was used as a proxy.

### *Agreement with the Gold Standard*

We calculated a number of standard metrics for agreement on R/E between the hospital discharge abstracts and each gold standard. Most important, we calculated accuracy (overall agreement), defined as the total number of individuals that are correctly assigned over the total number of individuals. Disagreement was defined as  $1 - \text{overall agreement}$ . We also calculated within category agreement for whites, non-whites, Hispanics, non-Hispanics, and non-Hispanic whites. These definitions were applied to the overall sample as

well as within each hospital. We also calculated sensitivity and specificity for identification of each R/E category. Agreement for “other” and for “unknown” has much less meaning as it is unclear what standards are used at individual hospitals.

We focused on the combined race/ethnicity measure. From a practical standpoint, all of the HCUP datasets report this measure. Furthermore, an artifact of the current racial and ethnic definitions has led to a high percentage of individuals of Hispanic background having either “other” or “unknown” race due to their mixed indigenous and other heritage. The inconsistency in racial identification is mitigated in the combined race/ethnicity measure, where known Hispanic ethnicity takes precedence over “other” or “unknown” race.

*Agreement with a Proxy Gold Standard: Hospital R/E Estimates*

Using the linked PDD-census data, we created, measured, and estimated population demographics by hospital. The proportion of patients of race  $j$  at hospital  $k$  is given as:

$$\text{prace}_{jk} = \frac{\sum_i \text{Race}_{ijk}}{\sum_{ij} \text{Race}_{ijk}} \tag{1}$$

While the estimated proportion of patients of race  $j$  at hospital  $k$  is given as:

$$\text{estprace}_{jk} = \frac{\sum_i \text{CensusRace}_{ijk}}{\sum_{ij} \text{CensusRace}_{ijk}} \tag{2}$$

where  $\text{CensusRace}_{ijk}$  is the proportion of individuals of race  $j$  living within the  $i^{\text{th}}$  patient’s zip code (at the  $k^{\text{th}}$  hospital). We created two disagreement metrics to compare the actual and estimated patient demographics. The absolute difference at the  $k^{\text{th}}$  hospital was calculated as:

$$\text{absdiff}_k = \frac{\sum_j |\text{prace}_{jk} - \text{estprace}_{jk}|}{2} \tag{3}$$

We also defined a similar metric, based upon the root mean squared difference (RMSD) between measured and estimated rates at the  $k^{\text{th}}$  hospital, given as:

$$\text{RMSD}_k = \frac{\sqrt{\sum_j (\text{prace}_{jk} - \text{estprace}_{jk})^2}}{\sqrt{2}} \quad (4)$$

Results of these two measures are comparable. The RMSD metric has a geometric (vector) interpretation that makes it somewhat more attractive than the absolute difference metric. These proxy gold standard metrics were created for all patients within a hospital as well as stratified by broad age categories within the hospital. Because measures were generally consistent regardless of stratification, we chose to retain metrics for the entire adult population. In these grand difference metrics, we performed sensitivity analyses including and excluding “unknown” as part of the error difference between the actual and estimated R/E distribution for the facility. Some facilities may report high rates of unknown race, but in reality, unknown race should occur relatively infrequently based upon reporting requirements.

We compared the hospital-level proxy gold standard agreement to the gold standard agreement within hospital for race/ethnicity reporting for the maternal cohort. As described, the cancer epidemiologic cohort had too few observations to make stable, consistent hospital estimates. We calculated correlation coefficients and simple fit to a line. We classified level of agreement by quartiles to identify whether certain hospitals (primarily low or high performers) could be reliably identified. We calculated missing and unknown race, ethnicity, and race/ethnicity by hospital. Hospitals were designated as outliers if their missing/unknown rate was greater than 10 percent. Hospitals were classified by quartile of missing/unknown. We then compared missing/unknown, quartile of missing/unknown, and outlier missing/unknown to the proxy gold standard measure of disagreement. Kappa coefficients were calculated based upon these comparisons. We also compared the quartiles for level of agreement to a number of hospital demographic measures from the gold standard data, including number of discharges (size), demographics (% white, Hispanic), gold standard mismatch rates (race—whites, race-nonwhites, race/ethnicity-overall, Hispanic, non-Hispanic), and completeness rates (unknown and missing).

To demonstrate the utility of the RMSD metric, we performed trend analysis between data from California and the six comparison states tracking the RMSD metric and rates of missing/unknown. For each state, we tracked average hospital RMSD and missing/unknown during the observation period. We also tracked the percent of outlier hospitals for these measures during



the period, flagging hospitals that either had greater than 10 percent missing/unknown and hospitals that had RMSD greater than 0.3.

The study protocol was reviewed and approved by the institutional review boards at the state and at UCLA. Data administration was performed using SAS 9.2 (SAS Institute, Cary, NC, USA), and analyses were performed using STATA 10 SE (STATA CORP, College Station, TX, USA).

## RESULTS

There were a total of 8 million cases in the 2008 and 2009 PDD. Of these, 1,052,238 hospital discharge summaries describing live births were linked to the state birth certificate registry. Of the 335 general acute care hospitals, 263 had linked deliveries in the maternal birth cohort. Two hospitals were dropped (one with a single observation and one with unreliable residence data). Using demographics from the PDD, adult patients discharged for deliveries were younger and more often Hispanic as compared to the cancer self-report cohort, which represents an older, non-Hispanic white population (Table 1). For the race-only measure, “unknown” race (mean 1.18 percent, range 0–19.8

Table 1: Demographics and Overall Agreement: Inpatient Data versus Maternal Cohort\* and versus Cancer Cohort†

	<i>Maternal Cohort 1</i>		<i>Cancer Cohort 2,3</i>	
	<i>N</i>	<i>% Gold Standard Agreement</i>	<i>N</i>	<i>% Gold Standard Agreement</i>
Total discharges	1,052,238		14,918	
Overall combined race/ethnicity		85.8		90.1
Overall race only		70.7		90.7
Discharges by race/ethnicity				
Hispanic	551,643	89.1	1,264	66.4
Non-Hispanic	500,595	82.1	13,654	92.3
Whites (all)	792,975	72.3	11,000	93.7
Non-whites	259,263	65.7	3,918	82.5
Mean age of patients at discharge (years)	28.1		62.8	
Number of hospitals involved	261		227	
Number of observations per hospital (mean, range)	4,001	<1, 15,263>	66	<1, 715>

\*PDD 2008–2009 versus Birth Cohort 2008–2009.

†PDD 2008–2009 versus Cancer Cohort: 2000–2008.

percent) and especially “other” race (mean 21.6 percent, range 0–95.4 percent) are reported quite frequently among the maternal cohort.

### *Overall Gold Standard Agreement*

Comparison of the maternal race/ethnicity data with the PDD index birth hospitalization reveals that overall correct identification was 86 percent (combined R/E) versus 71 percent (race alone). In the race-only measure, nearly 60 percent of individuals with “other” race and 50 percent with “unknown” race were Hispanic. Thus, the combined R/E measure generated higher correct identification through reclassification of both “other” and “unknown” to Hispanic, when possible. For the combined R/E measure, the positive predictive values (PPVs) were as follows: 0.78 for non-Hispanic whites, 0.96 for Hispanics, 0.86 for non-Hispanic Blacks, 0.92 for Asians, 0.43 for American Indian/Alaskan Natives (AIANs), and 0.045 for “other” race (Table 2). We also calculated sensitivity and specificity for identification of non-Hispanic whites (0.91 and 0.91), non-Hispanic blacks (0.91 and 0.99), Hispanics (0.89 and 0.96), and Asians (0.99 and 0.92).

In contrast, comparison of the PDD to self-reported data from the cancer cohort reveals a different picture. In this population, overall true-positive rate is 90 percent (Table 1). PPVs are 0.94 for non-Hispanic whites, 0.96 for non-Hispanic blacks, 0.70 for Hispanics, 0.93 for Asians, and 0.28 for AIANs (Table 2). We also calculated sensitivity and specificity for identification of non-Hispanic whites (0.95 and 0.87), non-Hispanic blacks (0.94 and 1.00), Hispanics (0.66 and 0.97), Asians (0.88 and 0.99), and AIANs (0.10 and 1.00).

### *Hospital-Level Gold Standard Agreement*

Hospitals were profiled based upon level of agreement. Among the 261 hospitals, the mean overall rate of disagreement for race was 29.6 percent and 15.4 percent for combined race/ethnicity. For race only, disagreement for whites was 27.6 percent and for non-whites 45 percent. For the combined race/ethnicity measure, disagreement for Hispanics and for non-Hispanics was 18.2 percent and 20.6 percent, respectively. Average rate of “unknown” race/ethnicity was 0.9 percent and “other” race/ethnicity was 2.7 percent. Thirteen hospitals in this sample were outliers (greater than 10 percent other/unknown). We compared overall agreement (excluding other and unknown) with missingness. The overall correlation between level of agreement and missingness was 0.34.

We compared the proxy measure—difference between reported and estimated race/ethnicity distribution by hospital (RMSD)—to the gold

Table 2: Gold Standard Comparisons: Inpatient Data versus Maternal Cohort\* and versus Cancer Cohort†

	TP	FN	FP	TN	SENS	SPEC	PPV	NPV	Acc	% Other	% Unknown	% Other and Unknown
Maternal cohort												
Non-Hispanic white	250,051	24,596	69,868	707,723	0.910	0.910	0.782	0.966	0.910	30.9	35.1	31.9
Hispanic	491,508	60,135	21,991	478,604	0.891	0.956	0.957	0.888	0.922	28.6	31.6	29.3
Non-Hispanic black	51,010	5,163	8,427	987,638	0.908	0.992	0.858	0.995	0.987	4.3	4.7	4.4
American Indian/ Alaskan Native	1,126	2,630	1,485	1,046,997	0.300	0.999	0.431	0.997	0.996	0.6	0.7	0.6
Asian Pacific Islander	107,501	22,389	8,954	913,394	0.828	0.990	0.923	0.976	0.970	28.8	22.1	27.3
Other	1,407	18,045	29,810	1,002,976	0.072	0.971	0.045	0.982	0.955	4.5	3.6	4.3
Unknown	206	16,471	8,894	1,026,667	0.012	0.991	0.023	0.984	0.976	2.2	2.3	2.2
Cancer cohort												
Non-Hispanic white	9,495	476	627	4,320	0.952	0.873	0.938	0.901	0.926	56.2	70.7	60.1
Hispanic	839	425	363	13,291	0.664	0.973	0.698	0.969	0.947	8.2	3.7	7.0
Non-Hispanic black	1,282	76	59	13,501	0.944	0.996	0.956	0.994	0.991	2.7	4.9	3.3
American Indian/ Alaskan Native	5	47	13	14,853	0.096	0.999	0.278	0.997	0.996	0.0	0.0	0.0
Asian Pacific Islander	1,797	248	137	12,736	0.879	0.989	0.929	0.981	0.974	23.7	20.7	22.9
Other	20	208	199	14,491	0.088	0.986	0.091	0.986	0.973	9.1	0.0	6.6
Unknown	-	-	82	14,836		0.995	0.000	1.000	0.995	0.0	0.0	0.0

\*PDD 2008-2009 versus Birth Cohort 2008-2009.

†PDD 2008-2009 versus Cancer Cohort: 2000-2008.

Acc, accuracy = (TP + TN)/(TP + TN + FP + FN); FN, false negative; FP, false positive; NPV, negative predictive value; SENS, sensitivity; SPEC, specificity; TN, true negative; TP, true positive.

standard measure based upon agreement with maternal R/E reported in birth certificates described above. R/E mismatch between discharge and birth certificates measured among maternal records was moderately correlated with the RMSD measures (0.44). Classifying the RMSD measures by rank tercile versus maternal record mismatch rank terciles showed low level of agreement overall (Kappa = 0.1). However, classification of the top 20 percent versus lower 80 percent showed moderate agreement overall (Kappa = 0.47).

We compared the RMSD to measures derived from the gold standard comparison and to the gold standard demographics (Table 3). The average RMSD in the lowest quartile was 5.7 versus 22.4 in the highest quartile. Across the hospitals with deliveries, rates of disagreement (viz. mismatch) within the gold standard comparison for race/ethnicity strongly track those in the RMSD measure. Conversely, mismatch for race does not strongly track with the RMSD measure for race/ethnicity. Hospital volume and unknown race are not strongly correlated with the RMSD measure, although the rate of “other” does have a significant trend.

Table 3: Gold Standard Agreement for Maternal Cohort\* versus Proxy Measure by Hospital

	Mean	Quartile Root Mean Squared Difference—Actual versus Expected Race/Ethnicity				p
		1	2	3	4	
N(hospitals)	261	66	64	65	66	–
Root mean squared difference—actual versus expected race/ethnicity (mean %)	13.5	5.7	10.8	14.9	22.4	–
No. of discharges per hospital (mean)	4,001	4,463	3,661	4,515	3,451	0.43
% Discharges (race; white)	77.6	84.2	80.4	73.1	72.2	<0.01
% Discharges (Hispanic)	50.6	67.4	46.2	46.9	42.0	<0.01
% Mismatch (race; overall)	29.6	36.7	26.4	30.4	24.5	0.35
% Mismatch (race; whites)	27.6	37.2	22.6	29.7	20.1	0.29
% Mismatch (race; nonwhites)	45.3	45.4	51.8	42.4	42.1	0.16
% Mismatch (race/ethnicity)	15.4	10.0	14.9	15.6	21.5	<0.01
% Mismatch (Hispanic)	18.2	9.1	19.3	17.3	27.5	<0.01
% Mismatch (non-Hispanic)	20.6	24.7	19.1	19.8	19.0	0.01
% Unknown (race/ethnicity)	0.91	0.70	1.04	1.17	0.75	0.34
% Other (race/ethnicity)	2.68	1.87	2.41	2.93	3.56	0.01

\*PDD 2008–2009 versus Birth Cohort 2008–2009.

*Assessing Data Quality across States Using the Proxy Measure*

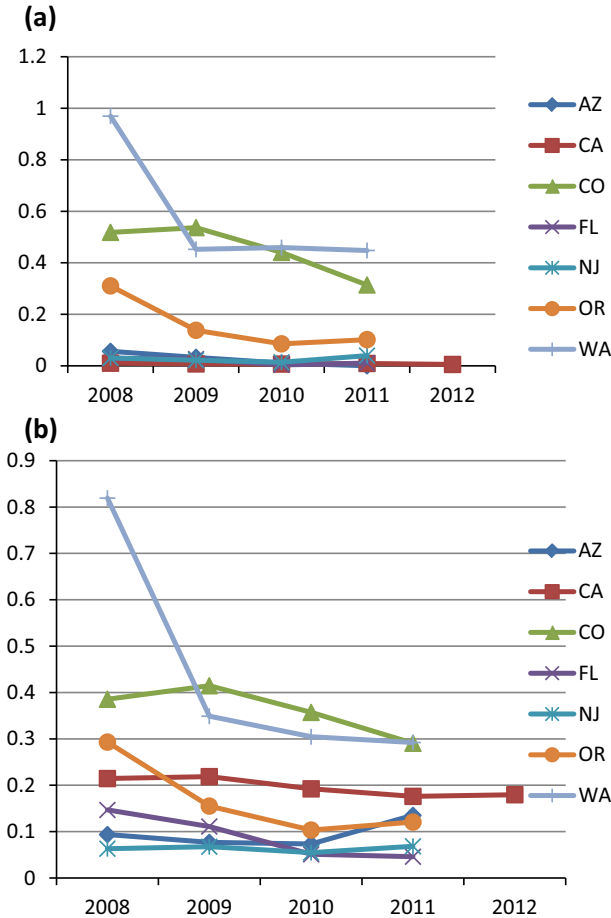
Data quality (agreement and completeness) varies across the comparison states. Application of the other/unknown measures plus the RMSD metric across the seven states showed general improvement between 2008 and 2011 for nearly all states (Figure 1a and b). Two states (CO and WA) had significantly worse RMSD measures (corresponding to 30+ percent disagreement) as recently as 2011 (Figure 1a). Three states (CO, WA, and OR) had much higher rates of combined “other” and “unknown” R/E than the other states, despite improvement between 2008 and 2011 (Figure 1b). High rates (>10 percent) of “other” and “unknown” R/E were common in many hospitals in WA (45 percent), CO (31 percent), and OR (10 percent) in 2011. High rates (>25 percent) of “other” R/E occurred commonly in many hospitals in CO (10 percent) as recently as 2011. Although we present the combined measure, the use of “other” and “unknown” was not consistently applied, as certain states preferentially used one category. For example, as recently as 2011, 40 percent of records in Washington were labeled “unknown” while Colorado continues to report approximately 10 percent of patients as “other.”

## DISCUSSION

Using existing statewide data from California and the Census, we examined the feasibility of creating proxy gold standard measures for assessing the accuracy of hospital reporting of R/E. The California hospital discharge data linked to the state birth registry enabled us to profile the accuracy of R/E reporting for mothers having babies by hospital. To a lesser extent, the much smaller cancer epidemiology cohort linked to the PDD highlighted the potential for age- and condition-related differences. Proxy overall agreement measures were more strongly correlated with hospital-level gold standard agreement for R/E reporting. When applied to track data quality, the RMSD measure demonstrated that certain states have begun to make data improvements, but overall, there has been little change in this measure across hospitals in recent years.

Statewide policies on the collection and auditing of R/E data by hospitals are evolving to improve completeness, consistency, and, to a much lesser extent, accuracy for an increasingly diverse patient population. Completeness approaches include tracking percent of records with other/unknown R/E. Consistency involves conforming to recognized racial and ethnic categories

Figure 1: Rates of High (>20 percent) Disagreement and High Combined Other and Unknown Race/Ethnicity by State. (a) Rates of High (>20 percent) Disagreement of Race/Ethnicity Reporting RMSD between Hospital and US Census. (b) Rates of High Combined Other and Unknown Race/Ethnicity Reported by Hospital



Notes. (a) Root mean square difference between average hospital reported race/ethnicity and that reported by the US Census. (b) Rates >10 percent by hospital.

(including the appropriate use of other/unknown), uniformity between datasets for an individual patient, implementing the correct procedures for combining multiple R/E, and for “rolling up” granular R/E information into the

major categories. Adopting consistency and completeness criteria and training staff to apply these to each patient are necessary first steps towards data improvement. Assessment of accuracy remains challenging due to the difficulty of obtaining gold standard comparisons.

Although a few states have adopted sophisticated data auditing standards, most states conduct only basic audit checks for data completeness and consistency. Among the comparison states in our study, Washington and Colorado currently have no formal data auditing procedures—in fact, R/E reporting in Colorado is voluntary. Oregon conducts basic data completeness and consistency checks. Florida audits for consistency and completeness and flags hospitals that report “unknown” race or ethnicity for more than 15 percent of records. Arizona has adopted a zero tolerance policy where hospitals are required to make corrections and resubmit their entire file if reported race does not match for patients appearing in multiple data sets within the last year. In addition, Arizona tolerates a 1.5 percent error rate and requires that race codes be consistent with recognized categories and no more than 4 percent of records report race as “refused.”

The nature of data reporting and review appear to be reflected in the levels of completeness and accuracy in the statewide data comparisons. States with no or basic auditing standards (CO, OR, and WA) showed worse agreement and had high percent of unknown compared to states that have more strict standards (AZ and FL). It is not surprising that states where R/E reporting is voluntary (CO) or has recently been mandated (WA) had higher rates of unknown and low rates of agreement. Although data completeness and accuracy differed by state, it is promising that improvements were made in all states over a short amount of time. Oregon recently adopted new standards and practices for collecting R/E data (Oregon Health Authority 2014), including an emphasis on collecting self-reported information. These new standards should lead to improvements in data completeness, consistency, and accuracy.

Initiatives have been started to improve data reporting and disseminate best practices. The New York State Department of Health is already using auditing measures developed through our work to reduce health disparities among Medicaid beneficiaries. They have created audit reports, based on hospital data linked to birth records and Medicaid enrollment that are reported back to hospitals. The success of their approach is based upon their ability to leverage access within the state bureaucracy to these other datasets and to link to them. Such data access and coordination varies by state. In states where the state hospital association collects and controls the hospital data, access to

Medicaid enrollment or birth records is more difficult. Access is not guaranteed even when the state bureaucracy controls all of these data. Linking data can prove difficult, time and resource intensive, and require navigating the policies of multiple data providers and overcoming administrative hurdles. While using self-reported data as the gold standard remains the best method of validating R/E data at the patient level, few sources of routinely collected gold standard data are available. In the absence of routinely collected self-reported data linked to hospital discharge data, metrics such as the RMSD provide the best possibility for assessing the accuracy of R/E data collected and feeding this information back to hospitals.

Accurate collection of R/E data can serve as a critical step in alleviating health disparities, improving quality, and informing policy (Institute of Medicine 2009). In a hospital setting, accurate collection of patient R/E may allow for identification of patient subgroups with unique needs and can inform the provision of resources and targeted interventions to improve quality of care, health outcomes, and, ultimately, alleviate health disparities. In the RWJF-funded *Expecting Success* program, 10 hospitals identified racial and ethnic disparities in heart failure outcomes, and then institutionalized the collection of self-reported patient R/E data to achieve improved core quality measures for patients admitted for heart attacks or heart failure (Robert Wood Johnson Foundation 2011). The Palo Alto Medical Foundation Research Institute utilized patient R/E collected in its outpatient facilities to elucidate prevalence of diabetic kidney disease among Asian subgroups (Bhalla et al. 2013), and racial and ethnic differences in cardiovascular risk factor control among minorities in northern California (Holland et al. 2013). These findings are informing tailored interventions for at-risk populations to eliminate these disparities. Accurate collection of R/E may also benefit hospitals financially, as more accurate risk adjustments can be made, translating into greater bonus payments. The Affordable Care Act has attempted to employ the new standards in a way that informs the national discussion on disparities by mandating these standards be consistently collected and reported in all national population health surveys relying on self-report data (The Patient Protection and Affordable Care Act 2010).

The study had a number of limitations. The validation study is based upon the analysis of data from a single state and two distinct cohorts, and may not be generalized to other populations or geographic areas. Our gold standard is based on women of childbearing age, which is different than other cohorts, such as older patients. We were unable to use a validated gold standard self-reported set of data. To our knowledge, the accuracy of the birth



certificate maternal R/E data has not been verified. Nevertheless, the birth certificate is meant to be completed by the birth mother or designated proxy, not by a hospital employee, and represents an exceptionally large source of self-reported demographic data. We linked cross-sectional data from the 2010 Census to hospital discharge data from a multiyear period, during which neighborhood and hospital characteristics may have changed. However, the hospital discharge data were collected within 1–2 years of the Census. Finally, there are discrepancies between the race categories used in the hospital discharge and birth records and Census data (which allow for selection of multiple races), which may adversely impact the agreement between hospital records and the population-based estimate.

## CONCLUSION

Identification of alternate measures for assessing reporting of R/E will be necessary as reporting requirements are enacted both federally and by individual states. Traditional measures are focused upon completeness, but not accuracy. Measures based upon comparisons between reported and estimated patient demographics appear to be reasonable first approximations for tracking hospital reporting, but they may not have enough precision for rank-order reporting. Differences in results by state are likely due to a combination of reporting practices and requirements as well as the underlying population mix. Comparison between reported and estimated hospital demographics appear to be reasonable for evaluating and tracking the overall accuracy of hospital reporting of R/E. Further work should focus on using this information to improve hospital collection of R/E data.

## ACKNOWLEDGMENTS

*Joint Acknowledgment/Disclosure Statement:* This study was supported by a grant from the Agency of Healthcare Research and Quality (1R01HS019963). The authors thank Roxanne Andrews at AHRQ and Jonathon Teague, Michael Kassis, and Joseph Parker at the CA Office of Statewide Health Planning and Development for their support of this work.

*Disclosure:* None.

*Disclaimer:* None.

## REFERENCES

- Agency for Healthcare Research and Quality. 2012. *National Healthcare Disparities Report 2011*. Rockville, MD: DHHS.
- Arday, S. L., D. R. Arday, S. Monroe, and J. Zhang. 2000. "HCFA's Racial and Ethnic Data: Current Accuracy and Recent Improvements." *Health Care Financial Review* 21 (4): 107–16.
- Bhalla, V., B. N. Zhao, K. M. J. Azar, E. J. Wang, S. Choi, E. C. Wong, S. P. Fortmann, and L. P. Palaniappan. 2013. "Racial/Ethnic Differences in the Prevalence of Proteinuric and Nonproteinuric Diabetic Kidney Disease." *Diabetes Care* 36 (5): 1215–21.
- Eicheldinger, C., and A. Bonito. 2008. "More Accurate Racial and Ethnic Codes for Medicare Administrative Data." *Health Care Financial Review* 29 (3): 27–42.
- Elliott, M. N., A. Fremont, P. A. Morrison, P. Pantoja, and N. Lurie. 2008. "A New Method for Estimating Race/Ethnicity and Associated Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity." *Health Services Research* 9: 69–83.
- Elliott, M. N., P. A. Morrison, A. Fremont, D. F. McCaffrey, P. Pantoja, and N. Lurie. 2009. "Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities." *Health Services Outcomes and Research Methods* 9: 69–83.
- Glaser, S. L., C. A. Clarke, S. L. Gomez, C. D. O'Malley, D. M. Purdie, and D. W. West. 2005. "Cancer Surveillance Research: A Vital Subdiscipline of Cancer Epidemiology." *Cancer Causes and Control* 16 (9): 1009–19.
- Gomez, S. L., and S. L. Glaser. 2005. "Quality of Cancer Registry Birthplace Data for Hispanics Living in the United States." *Cancer Causes and Control* 16 (6): 713–23.
- Gomez, S. L., J. L. Kelsey, S. L. Glaser, M. M. Lee, and S. Sidney. 2005. "Inconsistencies between Self-Reported Ethnicity and Ethnicity Recorded in a Health Maintenance Organization." *Annals of Epidemiology* 15 (1): 71–9.
- Hasnain-Wynia, R., and D. W. Baker. 2006. "Obtaining Data on Patient Race, Ethnicity, and Primary Language in Health Care Organizations: Current Challenges and Proposed Solutions." *Health Services Research* 41 (4 Pt 1): 1501–18.
- Holland, A. T., B. Zhao, E. C. Wong, S. E. Choi, N. D. Wong, and L. P. Palaniappan. 2013. "Racial/Ethnic Differences in Control of Cardiovascular Risk Factors among Type 2 Diabetes Patients in an Insured, Ambulatory Care Population." *Journal of Diabetes and Its Complications* 27 (1): 34–40.
- Institute of Medicine. 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care (full printed version)*. Washington, DC: The National Academies Press.
- Institute of Medicine. 2009. *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. Washington, DC: The National Academies Press.
- Kim, G., K. L. Ford, D. A. Chiriboga, and D. H. Sorkin. 2012. "Racial and Ethnic Disparities in Healthcare Use, Delayed Care, and Management of Diabetes Mellitus in Older Adults in California." *Journal of the American Geriatrics Society* 60 (12): 2319–25.

- Kim, G., C. B. Worley, R. S. Allen, L. Vinson, M. R. Crowther, P. Parmelee, and D. A. Chiriboga. 2011. "Vulnerability of Older Latino and Asian Immigrants with Limited English Proficiency." *Journal of the American Geriatrics Society* 59 (7): 1246–52.
- Kressin, N. R., B. H. Chang, A. Hendricks, and L. E. Kazis. 2003. "Agreement between Administrative Data and Patients' Self-Reports of Race/Ethnicity." *American Journal of Public Health* 93 (10): 1734–9.
- Morales, L. S., M. N. Elliott, R. Weech-Maldonado, K. L. Spritzer, and R. D. Hays. 2001. "Differences in CAHPS Adult Survey Reports and Ratings by Race and Ethnicity: An Analysis of the National CAHPS Benchmarking Data 1.0." *Health Services Research* 36 (3): 595–617.
- Morgan, R. O., I. I. Wei, and B. A. Virnig. 2004. "Improving Identification of Hispanic Males in Medicare: Use of Surname Matching." *Medical Care* 42 (8): 810–6.
- National Quality Forum. 2014. *Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors-Technical Report*. Washington DC: National Quality Forum.
- Office of Statewide Health Planning and Development. 2009. "Putting the Pieces Together: Linked Data Products" [accessed on May 1, 2014]. Available at [http://www.oshpd.ca.gov/HID/Data\\_Users\\_Conference/files/VitalStatsDeath.pdf](http://www.oshpd.ca.gov/HID/Data_Users_Conference/files/VitalStatsDeath.pdf)
- Oregon Health Authority. 2014. "Race, Ethnicity, Language, and Disability Demographic Data Collection Standards" [accessed on May 1, 2014]. Available at <http://www.oregon.gov/oha/oei/Documents/943-070%20Race%2c%20Ethnicity%2c%20Language%2c%20Disability%20Data%20Collection%20Standards.pdf>
- The Patient Protection and Affordable Care Act, P.L. 111-148, 23 March 2010.
- Robert Wood Johnson Foundation. 2011. *Expecting Success: Excellence in Cardiac Care: An RWJF National Program*. Princeton, NJ: Robert Wood Johnson Foundation.
- Stewart, S. L., K. C. Swallen, S. L. Glaser, P. L. Horn-Ross, and D. W. West. 1999. "Comparison of Methods for Classifying Hispanic Ethnicity in a Population-Based Cancer Registry." *American Journal of Epidemiology* 149 (11): 1063–71.
- Wei, I. I., B. A. Virnig, D. A. John, and R. O. Morgan. 2006. "Using a Spanish Surname Match to Improve Identification of Hispanic Women in Medicare Administrative Data." *Health Services Research* 41 (4 Pt 1): 1469–81.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.