

Agreement between Self-Reported and Administrative Race and Ethnicity Data among Medicaid Enrollees in Minnesota

Donna D. McAlpine, Timothy J. Beebe, Michael Davern, and Kathleen T. Call

Objective. This paper measures agreement between survey and administrative measures of race/ethnicity for Medicaid enrollees. Level of agreement and the demographic and health-related characteristics associated with misclassification on the administrative measure are examined.

Data Sources. Minnesota Medicaid enrollee files matched to self-report information from a telephone/mail survey of 4,902 enrollees conducted in 2003.

Study Design. Measures of agreement between the two measures of race/ethnicity are computed. Using logistic regression, we also assess whether misclassification of race/ethnicity on administrative files is associated with demographic factors, health status, health care utilization, or ratings of quality of health care.

Data Extraction. Race/ethnicity fields from administrative Medicaid files were extracted and merged with self-report data.

Principal Findings. The administrative data correctly classified 94 percent of cases on race/ethnicity. Persons who self-identified as Hispanic and those whose home language was English had the greater odds (compared with persons who self-identified as white and those whose home language was not English) of being misclassified in administrative data. Persons classified as unknown/other on administrative data were more likely to self-identify as white.

Conclusions. In this case study in Minnesota, researchers can be reasonably confident that the racial designations on Medicaid administrative data comport with how enrollees self-identify. Moreover, misclassification is not associated with common measures of health status, utilization, and ratings of quality of care. Further replication is recommended given variation in how race information is collected and coded by Medicaid agencies in different states.

Key Words. Race and ethnicity, measurement, Medicaid, administrative data

Racial and ethnic disparities in health care and health outcomes are major concerns for government agencies, academic researchers, and health care

practitioners. Addressing such disparities is a central goal of Healthy People 2010 (U.S. Department of Health and Human Services 2000). Successfully meeting this goal requires work at the state level particularly focused on individuals enrolled in public health programs such as Medicaid given that such programs provide services to a disproportionate share of minority populations in the United States (Centers for Medicare and Medicaid Services (CMS) 2000). Largely based on administrative claims data, academic researchers have recently documented important race disparities among Medicaid enrollees in a host of areas including the general use of health care services (Tai-Seale, Freund, and LoSasso 2001), behavioral health care (Opolka et al. 2003), and dental care (Dasanayake et al. 2002) as well as the treatment of AIDS (Kahn et al. 2002), cardiovascular disease (Litaker and Koroukian 2004), and diabetes (Shaya et al. 2005). Interpretation of these findings as well as the ability of states to both monitor and eliminate racial and ethnic disparities in the care provided by these programs rests heavily upon the collection of valid race and ethnicity information. Early work evaluating the utility of Medicaid data for health research (including demographic information) focused on the amount of missing observations and came to different conclusions about the validity of Medicaid data (Berkanovic 1974; Federspiel, Ray, and Schaffner 1976). There are no recent published evaluations of these data that take into account both the amount of missing data and the accuracy of records of individuals' race or ethnic identity compared with self-reports. There is, however, evidence suggestive of potential problems.

As part of the Balanced Budget Act of 1997, beginning in 1999 all states were mandated to submit their Medicaid claims data electronically to the CMS as part of the Medicaid Statistical Information System (MSIS). The specific "fields" or coding of information required was determined by CMS. Until 2005, states were asked to code race/ethnicity into nine categories: (1) white, (2) black or African American, (3) American Indian or Alaska Native, (4) Asian, (5) Hispanic or Latino (no race information available), (6) Native Hawaiian or Other Pacific Islander; (7) Hispanic or Latino and one or more races; (8) More than one race (Hispanic or Latino not indicated), and (9) Unknown. Since 2005, CMS only requires binary codes for each race/

Address correspondence to Donna D. McAlpine, Ph.D., Assistant Professor, School of Public Health, University of Minnesota, 420 Delaware Street SE, MMC 729, Minneapolis, MN 55455. Timothy J. Beebe, Ph.D., is with the Mayo Clinic College of Medicine, MN. Michael Davern, Ph.D., and Kathleen Thiede Call, Ph.D., are with the Division of Health Services Policy and Management, School of Public Health, University of Minnesota, Minneapolis, MN.

ethnicity (i.e., black/nonblack, etc.). While the states submit information to CMS that follow these codes, it is unclear how the information is initially collected and coded from Medicaid eligibility applications. As a recent report on the quality of data available to investigate race/ethnic disparities in health states: “CMS does not yet have any information on the quality of the racial and ethnic data collected through the MSIS” (National Research Council 2004, p. 92).

CMS details the many anomalies they have found in the MSIS data submitted by states (CMS 2005) which point to problems in the collection of race and ethnicity data at the state level. For example, in 2003 race/ethnicity was listed as “unknown” for more than one in five enrollees in Rhode Island, New York, and Vermont. Overall, in 2003 the field was unknown for approximately 6 percent of Medicaid eligibles.

A cursory examination of State Medicaid applications shows the diversity of methods for collecting information on race and ethnicity that may introduce measurement bias.¹ Some states (e.g., Alabama, Colorado, Iowa, and Illinois) include Hispanic as an option in the race question, instead of asking a separate question about ethnicity. Studies show that this may underestimate the number of Hispanics in the population (Beebe 2003). States such as Iowa, Massachusetts, and New York explicitly state that answering the questions about race and ethnicity are optional, which may impact the number of missing observations for these variables. In some instances, the application makes clear that the respondent may choose more than one race (e.g., Arizona, Pennsylvania, and Alaska), while in other states it appears to not be an option (e.g., Illinois, Colorado). Where standardized response options are provided, the format varies in at least three ways: check boxes, circle the answer, or insert a code from a prescribed list. States such as Massachusetts and Utah simply ask respondents to write in their race and ethnicity and provide no response options. Some states include an “other” category (e.g., Alabama, Colorado, Mississippi, and Hawaii) although it is not clear how they edit such information to fit within MSIS codes. This variability in question-asking is important because a long line of research in the survey literature has shown that responses can be affected in significant way by seemingly minor wording changes in the item stem and how response categories are cast (see Schuman and Presser 1981; Fowler 1995; Dillman 2000).

Findings from research examining the agreement between self-report and administrative measures of race and ethnicity have taken different approaches to the assessment of concordance. Many researchers have used

self-reports as the gold standard and look at the proportion of each group that are correctly classified in administrative data (the sensitivity of the administrative measure). The research suggests that the majority of individuals who self-identify as white or African American are also identified as such in administrative data. But that agreement is much lower for individuals who self-identify as Hispanic, Asian, or American Indian (Pan et al. 1999; Arday et al. 2000; Kressin et al. 2003; Gomez et al. 2005). Others have looked at who the administrative data is most likely to misclassify (the positive predictive value [PPV] of the administrative measure). Beal et al. (2006) report that administrative data (medical records) substantially also misclassify blacks, Asians, and American Indians. Studies that examine agreement between individuals' self-identification when asked open-ended questions find even less agreement between self-reports and administrative sources (Moscou et al. 2003; Buescher, Gizlice, and Jones-Vessey 2005). Indeed, as experiments conducted before the 2000 Census suggested individuals tend to conflate the concepts of race, ethnicity, and ancestry (Tucker et al. 1996) and do not always agree with the standardized response options such as white or African American that are typically provided in surveys.

Scant research, however, has investigated the characteristics of respondents (other than race or ethnicity) that are associated with misclassification on administrative data. A notable exception is research by Kressin et al. (2003) that utilized self-report and administrative data in a sample of Veteran Affairs enrollees. They found that disagreement between self-report and administrative measures was associated with having higher levels of education, more resources, and less utilization of health care. Moreover, persons with unknown race on administrative data were more likely to be married, young, better educated, and to have higher health status and lower utilization of services than those with known race on administrative data. Others have examined characteristics associated with misclassification on administrative data for persons self-identified as Asian, but found that demographic factors were not statistically significantly associated with misclassification (although sample size may have been too small to detect significant differences) (Gomez et al. 2005).

In this study, we present a case study of one state (Minnesota) to measure the agreement between survey and Medicaid data with respect to race and ethnicity. Two central questions are addressed: (1) Do measures of race/ethnicity from Medicaid administrative files agree with self-reported race/ethnicity? and (2) Are demographic, health status, or health care utilization characteristics associated with any observed misclassification?

METHODS

Data Sources

Medicaid Administrative Files. The Minnesota Department of Human Services (DHS) maintains administrative data on enrollees in their health care programs, including Medicaid (The Minnesota Medicaid Information System or MMIS). The project team obtained a file from DHS that contained all noninstitutionalized public health care program enrollees as of the date of the data were pulled (3/20/2003). The file contained information on each enrollee, including sociodemographic characteristics. For purposes of the current investigation, we focused on the data fields pertaining to enrollee race and ethnicity. These files are used to submit quarterly reports to CMS as part of MSIS.

In Minnesota, the system of data collection for race/ethnicity is largely decentralized and, at times, is the responsibility of the local Medicaid caseworkers. Although guidelines requiring caseworkers to obtain the race and ethnicity information via self-report at the point of application exist, it is unclear how well they are followed at the local level. The application includes check boxes for ethnicity (Hispanic and non-Hispanic) and race (five OMB categories), with ethnicity asked first. Respondents are advised that they can get help filling out the application from a local caseworker. Anecdotal evidence from persons working at DHS suggests that caseworkers are often uncomfortable asking about race and may try instead to “make a guess” through observation. There is no option for multiple race, and the form instructs “You do not have to answer the next two questions [ethnicity/race] if you do not want to.”

Survey Data. The self-reported racial and ethnic data come from a statewide survey of 4,902 Minnesota Health Care Program (MHCP) enrollees designed to assess racial and ethnic disparities in the use of preventive and other health services, as well as barriers to the use of these services. The sample consisted of a representative sample of all enrollees as well as oversamples of selected racial and ethnic groups. To understand barriers to service use for children and adults, the sample included randomly selected children and adult enrollees of the health care programs. When a child was chosen, an adult member of the household answered the questions about the child. The survey was conducted either by mail or by telephone between April and July 2003 and measured health status, health care utilization, and barriers to the use of services. The overall response rate was 54 percent (this is somewhat higher than rates typically seen in surveys of Medicaid enrollees where anything over 50 percent is considered very good) (Jensvold et al. 2003).

Measures

Respondents were first asked to identify their ethnicity, and then asked to identify their race; multiple responses to the race question were permitted. For purposes of analysis, the race and ethnicity responses are combined into a single variable; if the respondents indicated they were Hispanic, they were assigned that category on the race/ethnicity variable. Otherwise, they were assigned the value they answered for the race question. Owing to the low number of respondents who identified as "other" Pacific Islander ($N = 2$), this category was combined with Asian.

While there is no gold standard for the measurement of race and ethnicity; most agree that self-identity is superior to measures derived from other sources (Friedman et al. 2000; Mays et al. 2003). Thus, we use self-reported race/ethnicity as the gold standard to which the administrative measure is compared. We present data on the percent agreement between the two measures, as well as the sensitivity, specificity, and PPV. Sensitivity is the probability that administrative data correctly identified a specific race/ethnicity. Specificity is the probability that the administrative data correctly identified that a person was not of a specific race/ethnicity. The PPV is the probability of self-reporting a specific race/ethnicity among those identified as that race/ethnicity on the administrative data (or the percentage of persons defined on the administrative data as being a specific race/ethnicity who also so identified in the survey). We also calculate κ , a chance-corrected measure of agreement (Cohen 1960).

Measures of demographics and health experiences that are utilized in the analyses come from the survey data. These include global health, whether the respondent visited a doctor in the year before interview, overall rating of quality of care, age, home language (non-English versus English), marital status (married versus not married), and education status (less than high school versus high school or greater). When the sampled member was a child, the parental respondent answered the questions about the child's health, utilization, quality of care, age, gender, and home language. Questions about education status, marital status, and country of origin were asked of the sampled child's parent and thus measure attributes of the parent.

RESULTS

Table 1 provides the agreement between self-reported race and ethnicity and administrative data on race and ethnicity. Overall, agreement between the two

sources is high. Excluding those who were missing on the self-report ($N = 63$) or those who reported more than one race ($N = 225$), the administrative data correctly classified 94 percent of respondents.

Administrative data does not perform equally well for identifying each race/ethnic group. The sensitivity (shown in bolded row percentages) is lowest for those who self-identify as white (0.897) or Hispanic/Latino (0.909). For example, of those who self-identify as white on the survey, approximately 90 percent are also categorized as white in the administrative data. (The number of persons identifying as “other” race on the survey data are too small to interpret test statistics for this group.) Second, specificity is high (98 percent or higher) for each racial/ethnic category, indicating that there are few “false negatives” (e.g., the administrative data does well at identifying someone who is not African American as not African American).

Column percentages indicate the distribution of self-reported race/ethnicity within categories defined by the administrative data. Those in bold indicate the percentage of persons defined as a specific race/ethnicity on the administrative data who also self-report being from that group. In other words, the column diagonals indicate the PPV of the administrative measure of race/ethnicity. The PPV is lowest for American Indians and Hispanic/Latinos. For example, of the individuals that the administrative data classify as American Indian, 15 percent reported another racial/ethnic group in their survey responses (Table 1).

We next assessed whether concordant cases were significantly different than discordant cases, on race/ethnicity when controlling for other demographics, health status or use of services. This analysis is restricted to those who are classified in one of the five specific race/ethnicity in both the survey and administrative data. We excluded those who were missing/unknown on the administrative data because they are by definition discordant with the survey data. We excluded the “other” category because there are too few cases in the survey data to analyze this group.

Controlling for other demographics and health care experiences, self-reported race/ethnicity continues to be a significant predictor of discordance (Table 2). As shown, Hispanics/Latinos have over twice the odds of being classified as not Hispanic on the administrative data compared with whites. Persons who self-identify as African Americans and American Indians are significantly less likely than whites to be misclassified on the administrative data. This demonstrates that self-reported minority group membership is likely to be noted on the administrative data.²

With the exception of age, gender, and home language, none of the other coefficients reach statistical significance. It is interesting, however, that

Table 1: Agreement between Self-Reported Race/Ethnicity and Administrative Data on Race/Ethnicity ($n = 4,614$)

Survey Data	Administrative Data								N
	Hispanic/Latino Row % (Col %)	White Row % (Col %)	African American Row % (Col %)	Native American Row % (Col %)	Asian/Pacific Islander Row % (Col %)	Other Row % (Col %)	Missing/Unknown Row % (Col %)		
Hispanic/Latino	90.90 (94.89)	2.31 (1.34)	1.90 (1.07)	2.85 (4.30)	0.14 (0.13)	1.77 (21.31)	0.14 (2.78)	736	
White	1.88 (3.69)	89.73 (97.64)	1.16 (1.23)	2.39 (6.76)	0.22 (0.40)	2.53 (57.38)	2.10 (80.56)	1,383	
African American	0.46 (0.85)	0.38 (0.39)	96.64 (97.24)	1.30 (3.48)	0.23 (0.40)	0.76 (16.39)	0.23 (8.33)	1,311	
Native American	0.24 (0.14)	1.18 (0.39)	0.47 (0.15)	97.87 (84.84)	0.00 (0.00)	0.00 (0.00)	0.24 (2.78)	423	
Asian/Pacific Islander	0.26 (0.28)	0.40 (0.24)	0.40 (0.23)	0.40 (0.61)	97.89 (99.07)	0.40 (4.92)	0.26 (5.56)	759	
Other	50.00 (0.14)	0.00 (0.00)	50.00 (0.08)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	2	
N	705	1,271	1,303	488	750	61	36	4,614	
Specificity	0.99	0.99	0.99	0.98	0.99	NA	NA		
κ	0.92	0.91	0.96	0.90	0.98	NA	NA		

Notes: Specificity is the probability that the administrative data identified that the person was not a specific race/ethnicity; Sensitivity (bolded row percentages) is the probability that the administrative data identified that the person was a specific race/ethnicity; positive predictive value (bolded column percentages) is the probability a person self-reported a specific race/ethnicity, given the administrative data identified them as such; κ is chance corrected measure of agreement between source.

Table 2: Logistic Regression of Discordance between Administrative and Self-Reported Race/Ethnicity on Sociodemographics and Health Variables (1 = Discordant) ($N = 4,515$)[†]

	<i>N</i>	<i>OR (95% CI)</i>
Age		
< 18	1,724	Reference
18–64	2,291	0.70 (0.49, 0.99)*
65+	500	0.58 (0.29, 1.15)
Gender		
Male	1,817	Reference
Female	2,698	0.72 (0.53, 0.98)*
Race/ethnicity		
White	1,319	Reference
Hispanic	722	2.24 (1.45, 3.45)**
African American	1,298	0.58 (0.37, 0.91)*
American Indian	422	0.27 (0.13, 0.58)**
Asian/Pacific Islander	754	0.94 (0.45, 1.97)
High school graduate		
No	1,917	Reference
Yes	2,580	0.76 (0.52, 1.08)
Married		
No	2,448	Reference
Yes	2,038	0.78 (0.57, 1.09)
U.S. born		
No	1,966	Reference
Yes	2,544	1.32 (0.73, 2.42)
Home language English		
No	1,789	Reference
Yes	2,723	5.17 (2.63, 10.16)**
Health status		
Good/very good/excellent	3,635	Reference
Poor/fair	849	0.69 (0.43, 1.14)
Health care visit past year		
No	1,316	Reference
Yes	3,199	1.40 (0.96, 2.04)
Rating of health care		
Good/excellent	3,987	Reference
Poor/fair	497	0.83 (0.45, 1.53)

* $p < .05$.

** $p < .01$.

[†]Analyses do not include those who were missing race/ethnicity, “other race,” and those who answered multiple race.

CI, confidence interval.

persons who speak English as their home language have more than five times greater odds of having race/ethnicity misclassified on the administrative data than those who do not speak English at home.

We also assessed characteristics of respondents who were recorded as unknown or other race/ethnicity on the administrative data, but who did report a valid race/ethnicity on the survey data. We ran parallel analysis as shown in Table 2 (not tabled). Persons who self-identify as white are significantly more likely than each of the other racial/ethnic groups to be listed as unknown or other on the administrative data. This is not surprising as whites make up the majority of the Minnesota Medicaid population and as a result should make up the majority of the unknown or other cases as well. The only other social characteristics associated with being classified as other/unknown on the administrative data is marital status. Individuals who were married had more than double the odds of being recorded as unknown or other race/ethnicity on the administrative files compared with those who were not married.

Finally, we investigated characteristics of those who identified as being from multiple racial groups on the survey, an option not available in the administrative data. Approximately 40 percent of those who reported more than one race were classified as African American on the administrative data, while 36 percent were classified as American Indian. In multivariate analysis (not tabled), using the same independent variables in Table 2, but with classifications for race from the administrative data, we found that persons classified as African American (OR = 4.3), American Indian (OR = 5.4), or Asian/Pacific Islander (OR = 3.3) on the administrative data had significantly greater odds compared with those classified as white to report multiple races. The other social characteristics associated with reporting multiple races include home language; those who home language was English had 2.5 times higher odds than those who home language was not English of reporting multiple races on the survey. Child enrollees are also more likely than those who were 18–64 years of age to have multiple races reported on the survey data.

CONCLUSIONS

Studies have examined the quality of racial and ethnic designations in administrative data from specific states (Baumeister et al. 2000; Boehmer et al. 2002), in public health care programs such as Medicare (Pan et al. 1999; Waldo 2005), clinic records (Gomez et al. 2005) and surveillance systems for conditions such as cancer or AIDS (Kelly et al. 1996; Swallen et al. 1997). Few

have pursued such questions in a Medicaid population even though they are often the focus of health disparities research.

According to the results of the current investigation conducted in one state, we conclude that with respect to health status, health care utilization, and rating of health care quality there is not a difference in results dependent on whether we use the administrative data racial/ethnic measurement or the survey measurement. Our findings should make researchers reasonably confident that the Medicaid administrative data racial designations comport with how enrollees self-identify in Minnesota. The administrative data did have a slightly higher percentage of unknown or “other” race (2 percent) compared with the survey data (1.3 percent). The use of the “other” category is of most concern, because it is unclear what groups identify themselves as other, or if staff edit responses to exclude responses such as “American” or “Irish” which often occur when individuals report an “other” race. The level of missing information in these administrative data, however, is much less than has been reported in other states (CMS 2005). Moreover, the overall level of agreement is higher than has been reported in other states. For example, and in contrast to the current findings, Porter, Duncan, and Hu (2004) found *large* discrepancies between Medicaid enrollees’ self-reports of race and ethnicity and their racial and ethnic designations in administrative data in Florida. Further efforts are needed to better understand these divergent findings.

States’ ability to use these findings to improve their collection of race and ethnicity on eligibility files is somewhat limited, since we only have suggestive evidence to explain the high rates of concordance. We suspect that separate race and ethnic questions (consistent with OMB standards) and the use of check boxes corresponding to the five OMB categories of race helped improve concordance. We know little, however, about the process of collecting this information in other states. Because prospective enrollees have the option of filling out the form and supplying their racial and ethnic designations themselves, those designations have the potential to be self-reported in administrative data if they indeed respond to the items asking about race. If both the survey-based and administrative-based designations are self-reported, as such, there would be little reason to believe the two sources would diverge. As a point of comparison, a similar level of agreement was found in the Census 2000 reinterview study (Census households were reinterviewed with the same instrument after completing the original Census 2000 questionnaire) that found an overall 92.4 percent level of agreement in the edited race variables between the two surveys (Singer and Ennis 2002).

However, and also as indicated earlier, there is conjecture that third parties may be playing a role in the collection of race/ethnicity data on the enrollment form, especially if the applicant requires assistance from a county caseworker due to limited English proficiency. Moreover, some have suspected that in those instances where third parties are called upon to supply the applicant's racial and ethnic background, they ascertain that information via observation rather than asking, largely due to feelings of discomfort. Unfortunately, we cannot isolate those cases where someone other than the applicant supplied the racial and ethnic data with the available information. Future research ought to consider the specific issue of how often the classification of race in the administrative database may have been made by a third party (e.g., outreach worker, medical provider) observing the recipient's physical characteristics versus self-report.

While the overall level of agreement is excellent, the differences between administrative and self-reported race/ethnicity are not random. Persons who self-identify as Hispanic/Latino and those whose home language is English are more likely to be misclassified on administrative data compared to persons who identify as white or who report a language other than English as their home language. If these populations are the focus of investigation, administrative data may introduce error.

The administrative data also are more likely to assign more persons to minority group racial/ethnic group membership (Hispanic/Latino and American Indian) than self-report being from these groups. While the extent of the problem is small in the Minnesota data, it highlights a potential source of error. If administrative data misclassify whites into minority group status, analyses of disparities using these data may underestimate the true extent of differences between whites and racial and ethnic minorities.

An additional source of bias is cases that are missing on race/ethnicity on administrative data. Research on health disparities based solely on such data would delete these cases from analyses. Yet, the current investigation suggests that these enrollees are more likely to be white as whites make up the majority of population enrolled in Medicaid in Minnesota. Deleting missing cases, therefore, may introduce systematic bias into estimates of health status or health care utilization, a possibility requiring further study.

The overall response rate in the current study was 54 percent, raising the possibility that those who did not respond may be systematically different in some way that may effect the estimates of concordance. Analyses of persons who did not participate indicate that response rates were lowest in the stratum that oversampled minority racial and ethnic groups, and highest in the state-

wide simple random sample (mostly comprised of whites) (Beebe et al. 2003). The current study was also based on a stratified sample using administrative data classifications of race/ethnicity to create the strata. Oversamples were then drawn from minority group strata. The final sample, therefore, has a much higher representation of minority group members than the general population of Minnesota Medicaid enrollees. As a result our estimates of concordance could be biased to the extent that self-reported minority group members in the other/unknown and white group (all placed in the white stratum) are different from those identified as a minority group member in the administrative data.

Finally, the present findings are based on information from only one state. The generalizability of these findings can only be discerned via replication with the Medicaid administrative claims data from other states. Because the administration of Medicaid is left largely to states, the manner in which race data are collected, coded, and retained in states' administrative files is likely to vary. This variance is likely to bring about concomitant heterogeneity in the quality of administrative race and ethnicity data across states. Therefore, further state-specific replication of the current study is warranted and we intend to pursue this in future work. At the very least, those hoping to supplant self-reported race with those designations in Medicaid enrollment files in surveys of that population should do so with caution as those data may not comport with actual self-reports in certain states where concordance between the sources is low.

ACKNOWLEDGMENT

Partially funded by a grant from the Minnesota Department of Human Services.

NOTES

1. We examine Medicaid applications for 18 states; a list of states and detailed information about how they collect race and ethnicity information is available upon request from the corresponding author.
2. We also ran a model with the same universe of cases and dependent variable using administrative data race as an independent variable. This model shows the minority group membership on administrative data is a predictor of discordance meaning that people are more likely to be classified as a minority group member (American Indian or Hispanic) on administrative data than who self-report minority group race or ethnicity. These analyses are available from the authors upon request.

REFERENCES

- 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 Financing Review* 21 (4): 107–16.
- Baumeister, L., K. Marchi, M. Pearl, R. Williams, and P. Braveman. 2000. "The Validity of Information on 'Race' and 'Hispanic Ethnicity' in California Birth Certificate Data." *Health Services Research* 35 (4): 869–83.
- Beal, A. C., S. C. Chou, R. H. Palmer, M. A. Testa, C. Newman, and S. Ezhuthachan. 2006. "The Changing Face of Race: Risk Factors for Neonatal Hyperbilirubinemia." *Pediatrics* 117 (5): 1618–25.
- Beebe, T. J. 2003. *Measuring Race and Ethnicity in Health Insurance Surveys*. Minneapolis, MN: State Health Access Data Center.
- Beebe, T., H. Britt, T. K. Call, V. Cha, C. Kreider, J. Lundblad, D. D. McAlpine, J. McRae, B. Moore, S. Osman, and W. Suarez. 2003. *Disparities and Barriers to Utilization among Minnesota Health Care Program Enrollees*. Minneapolis, MN: Shadac, Division of Health Services Research and Policy.
- Berkanovic, E. 1974. "An Appraisal of Medicaid Records as a Data Source." *Medical Care* 12 (7): 590–5.
- Boehmer, U., N. R. Kressin, D. R. Berlowitz, C. L. Christiansen, L. E. Kazis, and J. A. Jones. 2002. "Self-Reported vs Administrative Race/Ethnicity Data and Study Results." *American Journal of Public Health* 92 (9): 1471–2.
- Buescher, P. A., Z. Gizlice, and K. A. Jones-Vessey. 2005. "Discrepancies between Published Data on Racial Classification and Self-Reported Race: Evidence from the 2002 North Carolina Live Birth Records." *Public Health Reports* 120: 393–8.
- Centers for Medicare and Medicaid Services. 2000. "A Profile of Medicaid: U.S. Department of Health and Human Services: Washington, DC" [accessed on April 28, 2005]. Available at <http://www.cms.hhs.gov/TheChartSeries/downloads/2Tchartbk.pdf>
- . 2005. "MSIS State Anomalies/Issues: All States: Department of Health and Human Services" [accessed on September 1, 2006]. Available at <http://www.cms.hhs.gov/MedicaidDataSourcesGenInfo/Downloads/anomalies.pdf#search=%22MSIS%20state%20anomalies%22>
- Cohen, J. 1960. "A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement." *Educational and Psychological Measurement* 20: 37–46.
- Dasanayake, A. P., Y. Li, S. Wadhawan, K. Kirk, J. Bronstein, and N. K. Childers. 2002. "Disparities in Dental Service Utilization among Alabama Medicaid Children." *Community Dentistry and Oral Epidemiology* 30 (5): 369–76.
- Dillman, D. A. 2000. *The Tailored Design Method*. New York: John Wiley & Sons Inc.
- Federspiel, C. F., W. A. Ray, and W. Schaffner. 1976. "Medicaid Records as a Valid Data Source: The Tennessee Experience." *Medical Care* 14 (2): 166–72.
- Fowler, F. J. 1995. *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage Publications.

- Friedman, D. J., B. B. Cohen, A. R. Averbach, and J. M. Norton. 2000. "Race/Ethnicity and OMB Directive 15: Implications for State Public Health Practice." *American Journal of Public Health* 90 (11): 1714–9.
- 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): 71–79.
- Jensvold, N. G., T. A. Lieu, F. W. Chi, A. M. Capra, and C. A. Sisk. 2003. "Strategies for Surveying Families of Medicaid-Insured Children by Telephone." *Journal of Health Care for the Poor and Underserved* 14 (1): 17–22.
- Kahn, J. G., X. Zhang, L. T. Cross, H. Palacio, G. S. Birkhead, and S. F. Morin. 2002. "Access to and Use of HIV Antiretroviral Therapy: Variation by Race/Ethnicity in Two Public Insurance Programs in the U.S." *Public Health Reports* 117 (3): 252,62; discussion 231–2.
- Kelly, J. J., S. Y. Chu, T. Diaz, L. S. Leary, and J. W. Buehler. 1996. "Race/Ethnicity Misclassification of Persons Reported with AIDS. The AIDS Mortality Project Group and the Supplement to HIV/AIDS Surveillance Project Group." *Ethnicity and Health* 1 (1): 87–94.
- 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.
- Litaker, D., and S. M. Koroukian. 2004. "Racial Differences in Lipid-Lowering Agent Use in Medicaid Patients with Cardiovascular Disease." *Medical Care* 42 (10): 1009–18.
- Mays, V. M., N. A. Ponce, D. L. Washington, and S. D. Cochran. 2003. "Classification of Race and Ethnicity: Implications for Public Health." *Annual Review of Public Health* 24: 83–110.
- Moscou, S., M. R. Anderson, J. B. Kaplan, and L. Valencia. 2003. "Validity of Racial/Ethnic Classifications in Medical Records Data: An Exploratory Study." *American Journal of Public Health* 93 (7): 1084–6.
- National Research Council. 2004. *Eliminating Health Disparities: Measurement and Data Needs*. Washington, DC: The National Academics Press.
- Opolka, J. L., K. L. Rascati, C. M. Brown, J. C. Barner, M. T. Johnsrud, and P. Gibson. 2003. "Ethnic Differences in Use of Antipsychotic Medication among Texas Medicaid Clients with Schizophrenia." *Journal of Clinical Psychiatry* 64 (6): 635–9.
- Pan, C. X., R. J. Glynn, H. Mogun, I. Choodnovskiy, and J. Avorn. 1999. "Definition of Race and Ethnicity in Older People in Medicare and Medicaid." *Journal of the American Geriatrics Society* 47 (6): 730–3.
- Porter, C. K., R. P. Duncan, and H. Hu. 2004. "Discrepancies in Race/Ethnicity between Medicaid Enrollees' Administrative Data and Survey Self-Report." Paper Presented at the 59th Annual Conference of the American Association for Public Opinion Research, Phoenix, Arizona, 2004.
- Schuman, H., and S. Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments in Question Form, Wording, and Context*. New York: Academic Press.
- Shaya, F. T., J. Y. Shin, D. Mullins, H. O. Fatodu, A. Gu, and E. Saunders. 2005. "Medicaid Managed Care: Disparities in the Use of Thiazolidinediones Compared with Metformin." *Journal of the National Medical Association* 97 (4): 493–7.

- Singer, P., and S. R. Ennis. 2002. *Census 2000 Content Reinterview Survey: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview, Census 2000 Evaluation B.5*. Washington, DC: U.S. Census Bureau.
- Swallen, K. C., D. W. West, S. L. Stewart, S. L. Glaser, and P. L. Horn-Ross. 1997. "Predictors of Misclassification of Hispanic Ethnicity in a Population-Based Cancer Registry." *Annals of Epidemiology* 7: 200–6.
- Tai-Seale, M., D. Freund, and A. LoSasso. 2001. "Racial Disparities in Service Use among Medicaid Beneficiaries after Mandatory Enrollment in Managed Care: A Difference-in-Differences Approach." *Inquiry* 38 (1): 49–59.
- Tucker, C., R. McKay, B. Kojetin, R. Harrison, M. de la Puente, L. Stinson, and E. Robison. 1996. "Testing Methods of Collecting Racial and Ethnic Information: Results of the Current Population Survey Supplement on Race and Ethnicity." Bureau of Labor Statistics.
- U.S. Department of Health and Human Services. 2000. *Healthy People 2010: Understanding and Improving Health*. Washington, DC: U.S. Government Printing Office.
- Waldo, D. R. 2005. "Accuracy and Bias of Race/Ethnicity Codes in the Medicare Enrollment Database." *Health Care Financing Review* 26 (2): 61–72.