The objective of this study was to assess how hospital designation of race/ethnicity based on Spanish surnames of mothers correlated with the self-report, and explores how these mothers identified their own race/ethnicity using the National Institutes of Health (NIH) categories. 235 mothers were enrolled prospectively and asked to report their race/ethnicity. Positive predictive value (PPV) of using surnames as a tool for assigning race/ethnicity was determined. The PPV of using surnames to identify ethnicity and race was 79 and 30%, respectively. 57% of mothers could not identify their race by the NIH categories. Although Spanish surnames more accurately reflected ethnicity than race, its use was associated with substantial discrepancies.

**Key words:** Race, ethnicity, self-report, Latino.

### INTRODUCTION

Determination of the race and ethnicity of enrolled subjects in clinical studies is important, as these characteristics may be associated with disease prevalence, risk factors, and outcomes. However, investigators often interchange these terms and may even consider them as synonymous. Historically, race has been defined as groups of people who have differences and similarities in biological traits, while ethnicity refers to those with a shared cultural heritage. As our knowledge of evolutionary forces such as genetic drift, founder effect, and selection improve, the distinction between race and ethnicity becomes increasingly important.

Government agencies in the United States historically have been inconsistent with its classifications of race and ethnicity. Starting in 1870, the U.S. Bureau of Census recognized five races: “White, Black, Mulattoes, Chinese and Indian.” In 1950, the categories were narrowed down to three: “White, Black and Other”. Forty years later, the United States Census broadened the racial categories back to five: “White, Black, Asian, Pacific Islander/Native Hawaiian, and American Indian/Alaska Native.” In 1990, “Hispanic” was added to the Census as an ethnicity. Currently, the National Institutes of Health (NIH) requires clinical investigators to classify a study population into two ethnic categories (Hispanic and Non-Hispanic) and six racial categories (American Indian/Alaskan Native, Asian, Black/African American, Native Hawaiian/Pacific Islander, White and other). Collecting race and ethnicity data for research purposes can be challenging as self-report is not always available. Surname lists were developed and are used by the United States Bureau of the Census as a method for identifying the Hispanic population. The use of surnames has since then been adopted by researchers and institutions in order to aid with classification of race and ethnicity. More information on the accuracy of such classification would be beneficial to researchers interested in ethnic and racial disparities.

As part of an observational study of very low birth weight (VLBW) infants admitted to the neonatal intensive care unit (NICU), mothers with Spanish surnames were asked about their ethnicity and race. The objective of this study was to assess whether the hospital assigned race and ethnicity based on Spanish surnames would match the self-reported race and ethnicity. Additionally, this study explored how mothers with Spanish surnames would identify their own race and ethnicity using the
METHODOLOGY

This prospective study enrolled mothers of VLBW infants (birth weight \( \leq 1500 \) g) who participated in the National Institute of Child Health and Human Development Neonatal Research Network’s (NICHD NRN’s) study entitled ‘The Generic Data Base Registry: Survey of Morbidity and Mortality in High Risk Preterm Infants’ from 1/1/2007 to 1/1/2010 at Parkland Memorial Hospital (PMH). This study provides a registry of baseline and outcome data for high-risk VLBW preterm infants, based on data collected in a uniform manner from NICUs at institutions participating in the NRN (Stoll et al., 2010). Specifically, mothers with Spanish surnames who agreed to have their infants participate in the study were verbally questioned. Trained research assistants questioned these mothers in person, with their preferred language, regarding their race and ethnicity. The Institutional Review Board of the University of Texas Southwestern Medical Center approved this study.

Mothers with Spanish surnames are considered by PMH as being Hispanic and white and are coded as such in the medical record based on their surname during registration. Mothers were questioned in person during their hospitalization. This study employed a set of questions based on the NIH race and ethnicity criteria for collecting this information on mothers with Spanish surnames. Specifically, the mothers were asked if they considered their ethnicity as being Hispanic. If they answered yes, then they were asked if they considered their race to be American Indian/Alaskan Native, Asian, Black/African American, Native Hawaiian/Pacific Islander, White or other. If they did not regard themselves as being Hispanic, then they were asked to state what ethnicity they considered themselves to be, followed by a designation of their race as stated earlier.

Statistical analysis

Data analysis was performed using Excel (Microsoft, Redmond, WA). Chi-square tests were used where appropriate. Using self-identification as the gold standard, positive predictive values (PPV) for using Spanish surnames as a classification tool for race and ethnicity was calculated. Comprehension of survey items was then assessed using response error analysis. A response error was defined by answers that were not consistent with a NIH category or were contradictory (that is, denying being of Hispanic ethnicity despite reporting their country of origin as Mexico). A p-value of \(<0.05\) was considered to be statistically significant.

RESULTS AND DISCUSSION

From 1/1/07 to 1/1/10, 235 mothers with Spanish surnames who delivered VLBW infants at PMH were enrolled. This represented >99% of eligible subjects. All of the infants were admitted to the PMH NICU. In the medical record, all of the mothers were classified as being of Hispanic ethnicity and white race based on their surnames. When the 235 hospital-defined Hispanic ethnicity mothers were questioned, only 187 (80%) regarded themselves as being Hispanic. Self-classification as Hispanic was significantly less than the hospital classification in which 100% of mothers were regarded as being of Hispanic ethnicity (\(p<0.05\)). Of the 48 (20%) mothers who did not consider themselves as being of Hispanic ethnicity, 40 (83%) reported “Mexican” as their ethnicity, 6 (13%) reported “Spanish” as their ethnicity, while the remaining 2 reported being “Black” and “American” (Figure 1).

When questioned regarding their race, 57% (134/235) of mothers could not identify with any of the NIH racial categories. Instead, they reported “Hispanic race” (83/235; 35%), “Mexican race” (37/235; 16%), “Indian race” (3/235; 1%), “Aztec race” (1/235; 0.4%), “natural race” (1/235; 0.4%) and “Spanish race” (1/235; 0.4%), and unknown race (8/235; 3%) (Figure 2). This is in contrast to all of them being classified as white by the hospital. Using self-identification as the gold standard, the PPV of using Spanish surnames as the tool for identifying Hispanic ethnicity was 80% in contrast to only 31% for race.

Determination of response errors was used for answers that were not consistent with NIH categories or contradictory (that is, denying being of Hispanic ethnicity despite reporting their country of origin as Mexico). Overall, there was a 38% (178/470) response error rate to survey items. Of the 235 answers to the question of ethnicity, 47 (20%) of the responses were contradictory in that mothers denied being of Hispanic ethnicity, but reported Mexican (83%) and Spanish (17%) origins, neither of which are recognized by NIH as an ethnicity. Of the 235 answers to the question of race, 131 (56%) answers were not consistent with the NIH racial categories and therefore were considered to be erroneous.

Accurate determination of the race and ethnicity of subjects enrolled in clinical and observational trials is important in order to define optimally their possible association with the condition studied. Such classification, however, has been a difficult and imprecise task (Gomez et al., 2005; Swallen et al., 1997; Hahn, 1992). In 1990, analysis of the national census data showed a sensitivity of 79% and a specificity of 90% when using Spanish surnames to predict the self-reported ethnicity. In 2003, a study conducted in a Veterans Administration Health System demonstrated that the race and ethnicity assigned by the hospital using surnames corresponded with patients’ self-report status in only 60% of the time (Kressin et al., 2003). Stewart et al., (2009) proposed that in the Hispanic population, when assessing ethnicity alone, classification based on medical record or surname represented a positive predictive value of 77 and 70%, respectively. Despite this, hospitals and researchers continue to use surnames to classify their population into specific ethnic categories (Wong et al., 2010; Perez-Stable et al., 1995; Morgan et al., 2004). The results of our study, where mothers with Spanish surnames identified themselves as Hispanic in only 80% of the cases, confirm these findings.

Similarly, race is a useful means of identifying specific populations that may possess unique risks for disease and thus require specific interventions. The importance of correctly classifying the race distinctly from ethnicity is highlighted by a few studies. Raskin et al. (1993, 1999)
showed that certain cystic fibrosis genetic mutations normally associated with White race was seen at high rates among Brazilians of lighter skin color and of Northern European descent. This stresses the importance of tracking genetic traits that can be associated with certain races. In contrast, O'Connor et al. (2002) showed that being of Hispanic ethnicity was associated with increased mortality rates in cystic fibrosis patients when compared with non-Hispanic individuals. This highlights the importance of understanding that certain ethnic groups can have a worse prognosis attributable to their cultural heritage rather than their genetic makeup. There are no studies that have assessed the accuracy of determining race apart from ethnicity among individuals with Spanish surnames. Our study suggests that the practice of using Spanish surnames to identify race correlated poorly with maternal self-report, with only 31% of mothers classifying themselves as belonging to a NIH-provided racial category.

Our study also showed that ascertaining ethnic and racial information from patients themselves did not improve the accuracy of their determination. This is reflected by the 38% response error rate to the questions of race and ethnicity. Mothers often reported their race to be Mexican and yet denied being of Hispanic ethnicity, suggesting a lack of understanding and poor identification of these categories among some mothers with Spanish surnames. This is consistent with the findings of a
previous qualitative study that showed poor identification with racial categories (Smith et al., 2010) and other studies that showed that particularly among the Hispanic population, the question of race is confusing and therefore often misreported (Mckenney and Cresce, 1993). These findings also correlate with data collected on the 1990 census whereby 40% of people who reported being of Hispanic ethnicity were not able to identify with one of the races provided and thus answered "other race" (Mckenney and Bennett 1994). It appears that the mothers in our study identified more strongly with their country of origin than the provided racial categories. Other reasons for the reduced accuracy among married women may be due to intermarriage and women adopting their husband’s surname.

A limitation of this study was that only mothers with Spanish surnames were enrolled, so it is possible that some Hispanic women with English surnames were not questioned. Moreover, we did not determine the educational level of the enrolled mothers since comprehension of the actual categories may have contributed to the responses. This was reduced by questioning the mothers in their preferred language, usually Spanish.

**Conclusion**

The use of Spanish surnames to identify the ethnicity and race of mothers who delivered VLBW infants at a public county hospital more accurately classified ethnicity than race, but substantial misclassification occurred. These data have important implications for studies that assess racial and ethnic disparities and health outcomes in this population. Research is needed on how best to ascertain race and ethnicity among these women and how their country of origin and education influence the self-designation of race and ethnicity.

**ABBREVIATIONS**

PMH, Parkland Memorial Hospital; NICU, Neonatal
Intensive Care Unit; NIH, National Institutes of Health; PPV, positive predictive value; NICHD, National Institute of Child Health & Human Development; NRN, neonatal research network; VLBW, very low birth weight.

REFERENCES


Smith FD, Woo M, Austin SB (2010). “I didn’t feel like any of those things were me’: results of a qualitative pilot study of race/ethnicity survey items with minority ethnic adolescents in the USA. Ethn Health. 6:621–38.


