The Concept of Race in Research: Using Composite Variables

Researchers who study differential outcomes based on racial classifications should acknowledge the sociopolitical forces that influence the concept of race. In this article, logistic regression findings based on a traditional methodological approach to race are compared to an approach that used a composite variable incorporating race/ethnicity and immigration status. Participants were 258 ethnically diverse low-income women drawn via convenience sampling from the Women, Infants, and Children’s Program (WIC) in or near a northeastern city. The PrimeMD Patient Health Questionnaire was used to identify whether the women had subthreshold or major depressive syndrome. The analyses using the composite variable better account for the findings than the analyses using race as a separate variable. Researchers should strive to identify and utilize various dimensions of participants’ social positions (e.g., immigration status, social economic status and language) that help explicate differential outcomes. (Ethn Dis. 2007;17:560–567)

Key Words: Race, Immigrant, Depression

The Concept of Race in Research: Using Composite Variables

As demographics shift and the racial and ethnic composition of the United States continues to build on the existing admixture of ancestry, researchers face methodological issues concerning racial and ethnic measurement. These categorizations, without biological substance, are the result of under- or overvaluation of particular groups of people who hold varying degrees of social capital. Devoid of any essentialist characteristics, the question of what accounts for the differences between various racial and ethnic groups and various health, economic, behavioral outcomes must be addressed. Defining the role and place for race in research is complex, involving both utility and conceptualization. A methodological approach that expands and appropriately identifies the myriad of dimensions that constitute the concept of race is needed.

Historically, race has been used to determine levels of difference and/or sameness between groups of people designated by racial categories. It has also been inappropriately used as a proxy for many macro-social forces, such as socioeconomic disparities between populations. Since 1790, the census has been the official designator of racial categories, long before modern genetic research and study. Federal racial categories have changed over time, corresponding with “demographic, political and ideological shifts in society.” The current categorization system from the Office of Management and Budget (OMB, 1997) modifies the system developed in the 1970s, which responded to the need to enforce civil rights laws and report compliance for groups that historically had experienced differential treatment because of race or ethnicity.

The current federal standards for data on race and ethnicity, established in 1997, include five racial categories: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White; two ethnic categories are also included: Hispanic or Latino, and Not Hispanic or Latino. OMB recommends that one or more responses to the question about race be allowed and that questions about race and ethnicity be asked separately, noting that in some circumstances, the two could be addressed together. In this case, “Hispanic or Latino” would be a sixth category and the variable would be better relabeled race/ethnicity.

Research funded by any governmental body must subscribe to the federal categorization system. Further, OMB recommends that these racial or race/ethnicity categories be used in household surveys, on administrative forms, and in medical and other research to provide consistent data on race and ethnicity. OMB acknowledges that these categories are “not anthropologically or scientifically based, but rather represent a social-political construct designed for collecting data on the race and ethnicity of broad population groups in this country.” (OMB, 1997)
Biological significance has proven to be an obscure predictor or correlate of particular variables and can limit the validity and reliability of studies that use racial taxonomies to explain differential outcomes between populations.\(^5\)\(^-\)\(^7\) Most agree that race is a social construction whose contribution rests on understanding and appropriately measuring the overall influence that ethnic and/or racial heritage has on an individual\(^8\)\(^,\)\(^9\) and that it is mutable, influenced by cultural zeitgeist and socio-political factors.\(^10\)

One tautological rationale for maintaining an essentialist categorization of race that many researchers assert is “real people use and recognize these classifications.”\(^11\) Others argue that since racism is real, disavowing racial categories would be tantamount to negating the fact that the consequences of racial oppression/discrimination have an effect on people’s health.\(^12\)\^-\(^15\) Having no better option and needing to explain the unequal distribution of health and wealth across social statuses, the categories of race remain. Until variables with more scientific validity can be created, we must understand the limits of race categories that we continue to utilize.

Research using racial classifications often addresses health disparities and the effects of racism/discrimination on individuals’ health.\(^16\)\^-\(^20\) Understanding differential access to healthcare services, health risk factors, and health status indicators is crucial to creating policies that address such inequities.\(^5\)\(^,\)\(^19\) Further, studies have examined the link between racial group differences and perceived discrimination, socioeconomic status, and other societal factors.\(^12\)\(^,\)\(^18\)\(^,\)\(^21\)\^-\(^23\) Increasing evidence exists that stress caused by societal forces such as racism, discrimination, and SES affect a person’s well-being.\(^18\)\(^,\)\(^21\)\(^,\)\(^23\) Increasingly, researchers are interested in tackling the methodological question of race, explaining the myriad meanings and constructs it holds as a variable.\(^24\)

**Capturing Complexity**

Researchers have attempted to capture the complexity of race/ethnicity by measuring the interactions between race, ethnicity, and variables such as socioeconomic status, country of origin and immigration status\(^5\)\(^,\)\(^18\)\(^,\)\(^20\)\(^,\)\(^21\)\(^,\)\(^23\) or by using these other classifications as covariates.\(^25\) Bennett explains that “data on ‘racial’ and ethnic differences should be analyzed in relation to potential confounding variables such as SES; and that the justification and methods for measurement of ‘race’ and ethnicity should be clearly defined and explicitly stated.”\(^5\) While this can add to a better understanding of how race and other social categorizations influence outcomes, the variable of race/ethnicity is still represented as one-dimensional. Others\(^24\) argue replacing race as a variable with various theoretical constructs (eg, racial identity, racism, stereotype threat, and social categorization) or combining race with other relevant variables to form composite variables\(^26\) would more appropriately comprise its complexity. In his research, Lazarsfeld\(^26\) examined the interchangeability of indices. If different indicators represent the same concept, they behave similarly mathematically. By implication, if the indicators or composite variables are not the same, their use will result in discrepant findings.

Further, the impact of variables may differ from population to population and study to study. Low-income Hispanic, Black, and White women who used public sector health care had no significant difference in rates of anxiety or affective disorders by race/ethnicity.\(^27\) Similarly, the National Comorbidity Study (NCS) found that the 12-month prevalence for any anxiety disorder was about the same for Whites, Blacks, and Hispanics.\(^28\) However, the Epidemiologic Catchment Area Study (ECA) found that the prevalence of specific anxiety disorders varied by race and ethnicity.\(^29\)\^-\(^31\) For mood disorders, the ECA found Whites had significantly higher lifetime prevalence than Blacks,\(^32\) while the NCS found Hispanics at greatest risk for affective disorders, followed by Whites and then Blacks.\(^28\)

**METHODS**

In this study, analyses using a composite variable that incorporated race/ethnicity and immigrant status, a variable salient to the study group, were compared to analyses using the more traditional, one-dimensional race/ethnicity categorization, either alone or in interaction with immigrant status. The construction of this particular composite variable does not mean that all composite variables should be constructed similarly. Examples of other factors that could add dimensionality to the variable race/ethnicity are: discrimination, nativity, citizenship, language, geography, and social economic status variables such as employment, income, and education. Other researchers are encouraged to identify factors of relevance to race/ethnicity in their particular studies and construct their composite variables accordingly.

This study is a secondary analysis of data collected for a study of mental health and service utilization of a large group of ethnically diverse women in the Women, Infants, and Children’s Program (WIC). WIC is a federal nutrition assistance program that provides specific types of foods monthly either directly or via vouchers, to low-income pregnant and postpartum women, infants and children less than five years of age who have or are at risk of diet-related problems.\(^33\) Of the 302 women enrolled in the larger study, 258 with the data needed for the comparative analyses were included in this study. Specific questions addressed in this study were:

1. Are differences in demographic characteristics of women in WIC, when categorized by race/ethnicity, comparable to differences between groups.

*Ethnicity & Disease, Volume 17, Summer 2007*
categorized by a composite variable constructed using race ethnicity and immigrant/migrant status?

2. Controlling for the effect of other covariates, is depressive symptomatology of women in WIC better accounted for by a) race/ethnicity alone; b) race/ethnicity and its interaction with immigrant/migrant status; or c) the composite variable?

Procedure and Sample
Participants were drawn via convenience sampling from WIC clinics in or near a northeastern city. Interviews were conducted between July and December 2003. Interviewers approached either all or a systematic sample of the women (eg, every other woman). Social work students in a graduate research class interviewed a fifth of the larger sample and two female research assistants, both fluent in Spanish, interviewed the remainder, all trained and supervised by the first author. Interviews took 20 to 30 minutes to complete and were conducted either before or after the participants’ WIC appointment at the WIC clinics. Interviewers helped respondents complete the survey as needed. Both English and Spanish versions of the survey were available. Those who completed the interview were given a $10 honorarium. The university human subjects review board approved the study. The overall sample size for the subjects reviewed was 562.

Measures
The Standard Federal Race/ethnicity Variable and the Composite Variable
Respondents described themselves as African American, Caribbean or West Indian, Asian, Latina-of-Puerto Rican descent, Latina-of-other descent, White or Caucasian American, White or Caucasian European, American Indian, or Other. Respondents also indicated if they were born in mainland United States. If not, they were asked to identify country or territory of origin. In order to compare findings based on the standard federal race/ethnicity categorization, either by itself or in interaction with immigrant/migrant status, the data were recategorized as follows. White or Caucasian Americans and White or Caucasian Europeans were reclassified as White; Latinas-of-Puerto Rican descent and Latinas-of-other descent were reclassified as Latina; and African Americans, Caribbean or West Indians were reclassified as Black. To construct the composite variable, both the self-description (African American, Caribbean or West Indian, Asian, etc.) and immigrant/migrant status (a term used because of Puerto Rico’s US territorial status) were used to create the following categories. US born Whites, US mainland-born Puerto Ricans, island-born Puerto Ricans, immigrant Latinas, immigrant West Indians, US born African Americans, and Other which included immigrant Europeans and Africans, and non-immigrant West Indians and Central and South Americans, none of which were present in sufficient number (10 or more) to create a separate category and differed enough from the other composite variable categorizations to preclude combining.

Demographic variables and covariates
The demographic variables and covariates included in this study were either identified by the literature or by one of the authors as relevant to the study’s outcome variables and include variables such as age, partner/marital status, whether someone helps with the childcare.

Outcome variables - depressive syndromes
The nine-item depression subscale of the PrimeMD Patient Health Questionnaire (PHQ) was used to identify women with indicators of a subthreshold depressive syndrome (Y/N). The PHQ is a screening tool used to detect the presence of probable psychological problems. The PrimeMD-PHQ is based on DSM-IV diagnostic criteria and was developed for use in primary care practice settings. The instrument is brief and highly reliable for diagnosing psychiatric symptoms. Spitzer et al found good agreement between PHQ diagnoses and those of independent mental health professionals (for the diagnosis of one or more PHQ disorders, k = 0.65; overall accuracy was 85%; sensitivity was 75% and specificity was 90%) in primary care settings. A PHQ-9 (the nine depression questions) score of >10 had a sensitivity and specificity of 88% for major depression. In a study of 3000 ethnically diverse obstetric-gynecologic patients, those with a PHQ diagnosis had significantly more functional impairment, disability days, healthcare use, and psychosocial stressors. The likelihood of having one or more PHQ diagnosis did not vary with ethnicity or whether the English or Spanish version of the tool was used. Further, the PHQ was tested among 1000 general hospital Spanish inpatients and performed well (k = 0.74; overall accuracy was 88%; sensitivity was 87%, and specificity was 88%).

Statistical Analysis
Descriptive statistics, chi-square, and logistic regression (backward elimination method with standard criteria) were the primary analytic tools. Chi-square analyses were used to ascertain relationships between race/ethnicity, immigrant/migrant status, their interaction, the composite variable, and the demographic variables, and the two outcome variables. Three logistic regression models were run for each depression outcome variable. Model 1 included race/ethnicity (which the bivariate analyses did not find to be significantly related to either outcome variables) and the significant covariates. Model 2 included the race/ethnicity, immigrant/migrant status (which the
Table 1. Demographics and covariates for women in WIC by the standard federal race/ethnicity and the composite variable (sample size)

<table>
<thead>
<tr>
<th>Demographics and covariates</th>
<th>The Standard Federal Race/Ethnicity Variable</th>
<th>Composite Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suburban residence</td>
<td>67.0</td>
<td>26.0</td>
</tr>
<tr>
<td>English not first language</td>
<td>0</td>
<td>36.0</td>
</tr>
<tr>
<td>18–25 years old</td>
<td>57.7</td>
<td>57.9</td>
</tr>
<tr>
<td>Single, div, or sep</td>
<td>66.0</td>
<td>63.6</td>
</tr>
<tr>
<td>Less than high school</td>
<td>19.6</td>
<td>37.7</td>
</tr>
<tr>
<td>Not working</td>
<td>64.9</td>
<td>55.3</td>
</tr>
<tr>
<td>No health insurance</td>
<td>4.1</td>
<td>11.8</td>
</tr>
<tr>
<td>Didn’t breastfeed†</td>
<td>51.1</td>
<td>39.4</td>
</tr>
<tr>
<td>Endorsed Western beliefs</td>
<td>61.1</td>
<td>50.0</td>
</tr>
<tr>
<td>Didn’t endorse traditional beliefs</td>
<td>53.7</td>
<td>56.0</td>
</tr>
</tbody>
</table>

The variables – family employment, number of children, age of youngest child, whether someone helped with childcare, and pregnancy status – were not included in the table as the differences were not significant.

* P < .1.
† P < .05.
‡ P < .01.
$ The sample for this analysis was smaller (n = 113) as only those with a baby one year or less in age were asked this question.

RESULTS

To answer research question 1, demographic variables that were differentially distributed by race/ethnicity or the composite variable are listed in Table 1. Variables that were differentially distributed by both were residence, language, age, education level, health insurance status, and whether Western beliefs were endorsed. Employment status and whether the respondent endorsed traditional beliefs were as differentially distributed by race/ethnicity only. A subset of women who had had a baby within the last year indicated whether they breastfed. Breast-feeding status was differentially distributed by race/ethnicity. Marital/partner status was differentially distributed by the composite variable only.

The variables – number of children, age of youngest child, whether someone helped the respondent with childcare, pregnancy status, and work in the family – were not differentially distributed by either race/ethnicity or the composite variable. Overall, 26.1% had three or more children, 47.9% had a child who was <1 year of age, 15.4% had no one to help with childcare, 18.3% were pregnant, and 27.6% had no work in the family. (Table 1) To answer research question 2, bivariate analyses were conducted to determine those variables that were significantly differentially distributed with the two outcome variables to inform the regression models. The variables differentially distributed with subthreshold depressive syndrome were the composite variable, marital/partner status, and work status. The variables differentially distributed with major depressive syndrome were race/ethnicity in interaction with immigrant/migrant status, marital/partner status, educational level, family employment, number of children and whether there was someone who could help the respondent with childcare. Neither race/ethnicity nor immigrant/migrant status was significantly related to either outcome variable. The interaction of the two was significantly related to major depressive syndrome. The composite variable was significantly related to both outcome variables.

For subthreshold depressive syndrome, the three logistic regression
models concurred that the odds for women who were not working were lower than those for working women (0.4, 0.4 and 0.3 respectively), and the odds for those who were single, divorced or separated were at greater risk than those for women who were married or partnered (3.7, 4.4 and 3.9 respectively). However, whereas Model 1 found no significant difference by race; and Model 2 found no significant difference by race, immigration status, or the interaction between the two; Model 3 did find significant differences between categories of the composite variable. Specifically, the odds for island born Puerto Rican women were significantly greater (5.6) than for White women. Further, Model 3 better accounted for the data than Models 1 and 2, as indicated by its higher $R^2$, model $\chi^2$, and lower $-2$ log likelihood and AIC values (Table 2).

For major depressive syndrome, the three logistic regression models concurred that the odds for women with: 1) less than a high school education were greater than those for women with more education (3.2, 3.7 and 3.8 respectively); 2) three or more children were greater than those with fewer children (3.1, 3.0, and 3.0); and 3) no one to help with childcare were greater than those who had someone to help (3.2, 3.8 and 3.5 respectively). Model 2 also found a significant interaction between the Black race/ethnic group and immigration status, meaning that the odds for major depressive syndrome for non-immigrant Blacks were much greater than those for Black immigrants. Model 3 found the odds for African Americans to be significantly greater than those of Whites. All models accounted for the data similarly (Table 2).

**DISCUSSION**

The overall goal of this study was to evaluate how analyses using a composite variable that fleshed out the variable of race/ethnicity compared to analyses based on race/ethnicity or the interaction between race/ethnicity and immigrant/migrant status. Descriptive findings and risk profiles derived from logistic regression were compared. The findings indicated that very different pictures emerge in both cases. For the descriptive analyses, many differences between the Puerto Rican and Latina subgroups and the Black subgroups...
If the logistic analyses had only included race/ethnicity, the relevance of race/ethnicity in concert with immigrant/migrant status would have gone undetected. Would have been lost without the composite variable. If the logistic analyses had only included race/ethnicity, the relevance of race/ethnicity in concert with immigrant/migrant status would have gone undetected.

Strengths and Limitations

The use of the composite variable revealed important and informative group differences, illuminating the inadequacy of the usual race/ethnicity variable. However, statistically many of the analyses lacked power, as cell size was small. Although the PHQ has been validated among Spanish speaking samples, the extent to which these findings can be generalized to this study’s specific sample is unknown and its utility for immigrant West Indians untested. In addition, data on variables that could further flesh out race/ethnicity, such as discrimination, were not collected. Many empirical studies that demonstrate differential health/mental health outcomes between dominant and subordinate racial/ethnic groups have implied that such differentials are a consequence of perceived discrimination, but most research has not measured discrimination as a separate variable. Socioeconomic status variables such as income, occupation, and education are more commonly identified as the social causes of health/mental health disparities, but discrimination may be antecedent to such resource-related positions. A few studies have included discrimination measures and tested for associations between discrimination and health mental health and mental health utilization. Each study demonstrated how certain aspects of discrimination played a differential role in an individual’s experience. Understanding the social context of racism and discrimination, many researchers tend to speculate on associations to the differential outcomes, but do not provide tangible measurement of the relationship between perceived discrimination and participant outcomes. One drawback of including measurements of discrimination and/or racism is that they generally tend to be based on participant self-report. Yet the shortcomings are outweighed by the importance of fleshing out salient components of race/ethnicity in analyses.

Acculturation is another process that has been shown to affect immigrant health. However, similar to the issues surrounding the measurement of discrimination, acculturation instruments have been scrutinized for being unidimensional, the subject of less than robust analysis, and harboring unexamined latent variables. Most means of measurement focus on the individual and his/her particular behaviors, values, rituals, or norms and fail to include important structural and contextual factors, such as racism, discrimination, and segregation, which have an enormous effect on how and to what an individual acculturates. More importantly, studies that have attempted to understand the effects of acculturation on health have found the relationship to be complex and multi-faceted. Some studies have concluded that more acculturation (into dominant US American culture) has shown negative risk factors for health, while others have associated it with healthy behaviors. More research is needed to better understand acculturation and how it influences attitude, perception and behavior towards health. The research is best approached by a mixed methodological agenda beginning with qualitative studies that create an evidence-based model that represent salient individual, contextual and structural concepts and followed by quantitative testing of the model.

Implications

The concept of race and the use of the federal standard race/ethnic categorization are crucial to ongoing research about disparities in health and mental health between various populations. However, using socially prescribed categories that are constructed by socio-political forces reduces the meaning of what aspects of a population are really being measured and reifies racial/ethnic stereotypes. Yet applying “color-blind” methodologies negates the reality that social positions contribute to differential outcomes. Caution must also be used when substituting the variable of race with that of other social factors so as not to diminish awareness of how various forms of racism effect health disparities.

The argument for using composite variables does not mean that racial/ethnic classifications are always flawed, but rather that the fluidity of such a classification can restrict its usage as an independent variable. If and when utilized, researchers must be attentive to how such one-dimensional categorizations limit the scope of their investigation. Researchers must strive to identify and utilize factors that help explicate differential outcomes occurring as a result of being a member of a particular racial or ethnic social group. The heterogeneity that exists in various racial and ethnic social groups requires that variables be multi-dimensional and that instruments that measure socio-political forces such as racism and discrimination, factors that are difficult to quantify, be used as well. Correspondingly, analyses that compare findings based on the unidimensional race/ethnicity variable, to those based on its interaction with identified factors, to those based on a composite variable will...
further elucidate and disentangle the forces that underlay race/ethnicity. Examples of other factors that could add dimensionality to the variable race/ethnicity are: discrimination, nativity, citizenship, language, geography, and social economic status variables such as employment, income, and education. Researchers are encouraged to identify factors of relevance to race/ethnicity in their particular studies and construct their composite variables accordingly.

**ACKNOWLEDGMENTS**

The research was supported by a grant from the University of Connecticut. The authors thank the WIC women who participated in the study and WIC personnel, Dr. Ming-Hui Chen for statistical consultation, and Dr. Cheryl Parks for help with reviewing and editing.

**REFERENCES**


**OTHER SOURCES**

**AUTHOR CONTRIBUTIONS**
Design concept of study: Kurz, Werkmeister
Acquisition of data: Kurz
Data analysis and interpretation: Kurz, Werkmeister
Manuscript draft: Kurz, Werkmeister
Statistical expertise: Kurz