HHS Public Access

Author manuscript

Ethn Health. Author manuscript; available in PMC 2019 April 01.

Published in final edited form as:

Ethn Health. 2018 April ; 23(3): 233–248. doi:10.1080/13557858.2016.1263284.

Distinguishing between primary and secondary racial identification in analyses of health disparities of a multiracial population in Hawaii

Andrew Wey[a,b], **James Davis**[a], **Deborah Taira Juarez**[c], and **Tetine Sentell**[d]

[a]Office of Biostatistics and Quantitative Health Sciences, John A. Burns School of Medicine, University of Hawaii at Manoa, Honolulu, HI, USA

[b]Chronic Disease and Research Group, Minneapolis Medical Research Foundation, Minneapolis, MN, USA

[c]Department of Pharmacy Practice, The Daniel K. Inouye College of Pharmacy, Honolulu, HI, USA

[d]Office of Public Health Studies, University of Hawaii at Manoa, Honolulu, HI, USA

Abstract

Objective—To examine the importance of distinguishing between primary and secondary racial identification in analyzing health disparities in a multiracial population.

Methods—A cross-sectional analysis of 2012 Hawaii Behavioral Risk Factor Surveillance System (H-BRFSS). As part of the survey, respondents were asked to identify all their races, and then which race they considered to be their primary race. We introduce two analytic approaches to investigate the association between multiracial status and general health: (1) including two separate dichotomous variables for each racial group (primary and secondary race; for example, 'primary Native Hawaiian' and, separately, 'secondary Native Hawaiian'), and (2) including one combined variable for anyone choosing a particular racial group, whether as primary or secondary race ('combined race'; e.g. Native Hawaiian). Linear regression then compares the multiracial health disparities identified by the two approaches, adjusted for age and gender.

Results—The 2012 H-BRFSS had 7582 respondents. The four most common self-identified primary racial/ethnic groups were White, Japanese, Filipino, and Native Hawaiian. Native Hawaiians were the largest multiracial group with over 80% self-identifying as multiracial. Health disparities for Native Hawaiians, Portuguese and Puerto Ricans were attenuated by 10% after accounting for multiracial status. Populations that self-identified secondarily as Japanese, Puerto Rican, Mexican, and other PI had significantly poorer self-reported health.

Conclusion—The analysis illustrates the importance of accounting for multiracial populations in health disparities research and demonstrates the ability of two approaches to identify multiracial

CONTACT Andrew Wey, awey@hawaii.edu, Office of Biostatistics and Quantitative Health Sciences, John A. Burns School of Medicine, University of Hawaii at Manoa, 651 Ilalo St., Suite 211, Honolulu, HI 96813, USA.

Disclosure statement

No potential conflict of interest was reported by the authors.

health disparities in data sets with limited sample sizes. The ‘primary and secondary race’ approach might work particularly well for a multicultural population like Hawaii.

Keywords

Health disparities; multiracial status; self-reported health; statistical methods

1. Introduction

Racial health disparities are well-documented in the United States (US) across a myriad of health outcomes, time frames, and locations, which motivates considerable research and policy action to understand and reduce these disparities. Population-based research on racial/ethnic health disparities almost uniformly ignores ‘multiracial’ populations despite strong evidence that health disparities exist in multiracial populations and the growing multiracial population in the US, representing an estimated 6.9% to 12% of the population. Race is instead considered as discrete categories and mixed race individuals are asked to select a primary racial group, included as ‘other race/ethnicity,’ analyzed as a single ‘mixed race’ group, or simply excluded from studies of racial health disparities. These current approaches fail to capture the true diversity of a population and may obscure critical disparities in multiracial populations.

These are important problems as identifying disparities in multiracial populations is critical to implementing effective policies to improve population health. Additionally, better understanding the relationship between multiracial identity and health status can help untangle the complicated pathways in which race is itself associated with health outcomes (Griffith et al. 2010), as well as the fact that race also stands as a proxy for other factors, such as poverty or residential segregation (Mays, Cochran, and Barnes 2007). Such clarity is necessary to fully understand social determinants of health and to build effective population-based health policies to address persistent racial health inequities in the US (Griffith et al. 2010). This issue is particularly relevant to the many US locations where the percent of those who identify as multiracial is considerably higher than the national average (Krogstad 2015). For example, nearly one in four residents in Hawaii self-identify as multiracial, which is the highest percentage in the US (Krogstad 2015; Patten 2015). Ignoring multiracial identity in population-based studies in these locations may mispresent health disparities and pathways and lead to incorrect policy recommendations.

Significant practical and methodological obstacles prevent the investigation of multiracial health disparities in most population-level data. In many cases, a study will only collect the single self-identified primary race for each respondent, which simplifies data collection, but precludes the investigation of multiracial health disparities (Bratter and Gorman 2011). Another common approach allows participants to self-identify more than one racial group. This is the current Office of Budget and Management standard (United States Department of Health and Human Services 2011) and the current approach implemented in the US Census (Humes, Jones, and Ramirez 2011). Yet practical challenges exist in the analysis of such data, particularly in obtaining sufficient sample sizes to investigate the resulting subpopulations (Bilheimer and Klein 2010). Additionally, as race is often socially

constructed as one specific identity (Bratter and Gorman 2011), the full racial mixture of an individual may not be represented by the primary racial/ethnic affiliation. Thus, participants who self-identify as more than one racial group are often additionally asked to self-identify the group that best represents their race, or their primary self-identified race. This third approach is now standard in many major health surveys, including the National Health Interview Survey and the Behavioral Risk Factor Surveillance System (BRFSS) (Anderson, Bulatao, and Cohen 2004). Analytic uncertainty exists concerning the relative importance of primary versus the secondary self-identification.

Another major practical issue facing multiracial health disparities research is the small sample sizes of multiracial populations in most population-based research. In practice, even large national health surveys do not have adequate sample sizes in multiracial populations and combining multiple survey years is commonly used to circumvent this problem. For example, Bratter and Gorman used seven years and over one million total respondents of the BRFSS to achieve a large enough sample size to investigate multiracial health disparities at the national level (Bratter and Gorman 2011). However, this approach is not feasible for state and local population-based studies due to limited sample sizes. Thus, we need an approach that can handle the potentially small sample sizes of multiracial populations.

The goal of our study was to address these research gaps by comparing different analytical approaches for including multiracial status in population-based research. This data will help researchers and policy-makers understand the relative importance of the distinction between primary and secondary racial self-identification in understanding health outcomes. It can also help illuminate the utility of distinct methods of collecting multiracial status, especially given the small sample sizes of multiracial populations in most population-based research.

Specifically, we provide a formal comparison of two analytic approaches for investigating the impact of multiracial status on health outcomes: (1) considering two separate variables for each racial group, one if included as a primary race and one if included as a secondary race (‘primary and secondary race’); and (2) considering one combined variable for anyone choosing a particular racial category, whether as primary or secondary race (‘combined race’). This assumes that the distinction between the primary and other, secondary, races is unnecessary. These two approaches and the approach that only considers the primary race (‘primary race only’) are compared to each other with the 2012 Hawaii Behavioral Risk Factor Surveillance System (H-BRFSS) data. We then consider both appropriateness of each analytical approach for handling limited sample sizes and distinguishing between primary and secondary racial self-identification.

Hawaii and the H-BRFSS provide an excellent illustration on the importance of accounting for multiracial health disparities as over 20% of Hawaii identifies as multiracial (Jones and Smith 2001), likely reflecting the future demographics of other locations in the US. In another distinct advantage, Hawaii also has wide diversity within its multiracial/ethnic population mix, including Asians, Pacific Islanders, Whites, and Hispanics in relatively high numbers. [More detail around these racial/ethnic populations in Hawaii can be found in the Appendix.] Thus, with the H-BRFSS, we are able to investigate multiracial health disparities

on self-reported general health (SRH) and compare methods distinguishing between the primary and secondary self-identified races with relevance to other population-based studies.

2. Approaches for investigating multiracial health disparities

For illustrative purposes, we introduce the specification and interpretation of the analytical approaches using only white and Native Hawaiian (NH) self-identification. The NH comparison is important due to known health disparities, a notably multiracial/ethnic population, and evidence of differential health disparities between self-identified full versus part NH (Kaneshiro et al. 2011).

2.1. The three analytical approaches

2.1.1. Primary race only—A common approach for modeling racial health disparities uses only the self-identified primary race, which is referred to as the 'primary race' approach through the rest of the paper. This approach *implicitly* assumes that multiracial populations possess the same health as the self-identified primary race. This has the advantage of simplified analyses, but fails to capture any variability in self-identified secondary races or multiracial populations. An unadjusted linear regression model for this approach for Native Hawaiians is straightforward and indicated by:

$$E\{Y|\mathrm{PR}_W, \mathrm{PR}_{\mathrm{NH}}\}=\alpha_0+\alpha_{\mathrm{NH}} \times \mathrm{PR}_{\mathrm{NH}}, \quad (1)$$

where Y is the random variable for SRH, PR_W and PR_{NH} are indicators for, respectively, primary white and NH self-identification. The interpretation of α_{NH} is the average difference in SRH between the populations that self-identify as NH versus self-identify as white.

2.1.2. Primary and secondary race—Another option, an extension of the 'primary race' option, includes indicator variables for the self-identified secondary races. This is referred to as the 'secondary race' approach throughout the rest of the paper. An unadjusted linear regression for SRH with the 'secondary race' approach is

$$E\{Y|\mathrm{PR}_W, \mathrm{PR}_{\mathrm{NH}}, \mathrm{SR}_W, \mathrm{SR}_{\mathrm{NH}}\}=\alpha_0+\alpha_{\mathrm{NH}} \times \mathrm{PR}_{\mathrm{NH}}+\gamma_W \times \mathrm{SR}_W+\gamma_{\mathrm{NH}} \times \mathrm{SR}_{\mathrm{NH}}, \quad (2)$$

where Y is the random variable for SRH, PR_W and PR_{NH} are indicators for, respectively, primary white and NH self-identification and SR_W and SR_{NH} are indicators for, respectively, secondary white and NH self-identification. The interpretation for α_{NH} is the average difference in SRH between the population that self-identifies primarily as NH compared to the population that self-identifies primarily as white given that secondary racial self-identification is the same. In contrast, the interpretation for γ_W is the average difference in SRH between self-identifying and not self-identifying secondarily as white given that the primary self-identification and other secondary self-identification is the same. Finally, the intercept, α_0, is the average SRH for whites with no secondary self-identification. Note that

the ‘secondary race’ approach assumes that the association of secondary self-identification is the same regardless of the primary race.

Equations 1 and 2 demonstrate that the common approach of using primary self-identified race is nested within the ‘secondary race’ approach. Thus, an advantage of the ‘secondary race’ approach is that self-identified primary race remains in the model and maintains a similar interpretation, while also controlling for multiracial identity. Despite the straightforward extension, the ‘secondary race’ approach may encounter issues with small sample sizes as some racial groups may have low secondary racial self-identification.

2.1.3. Combined race—To help resolve the sample size issue, note that the distinction between the self-identified primary and secondary races essentially splits the population that self-identifies as a given race into two separate populations. As such, a natural approach to alleviate the sample size concern might assume that the distinction between the primary and secondary self-identified races is unnecessary; see Figure 1 for an illustration of the racial self-identification process. This approach would include indicator variables for *all* self-identified races, and is referred to as the ‘all self-identified races’ approach throughout the rest of the paper. As an example, an unadjusted linear regression for SRH with the ‘all self-identified races’ approach is

$$E\{Y|\mathrm{AR}_W, \mathrm{AR}_{\mathrm{NH}}\}=\beta_0+\beta_W \times \mathrm{AR}_W+\beta_{\mathrm{NH}} \times \mathrm{AR}_{\mathrm{NH}}, \quad (3)$$

where Y is the random variable for SRH, and AR_W and AR_{NH} are indicators for *any* white and NH self-identification, respectively. This means that the populations estimating β_W and β_{NH} have individuals in common and the intercept, β_0, is similar to a weighted mean of SRH without differences between racial populations. The difference between white and NH self-identification, which corresponds to $\beta_W - \beta_{NH}$, possesses a subtle change of interpretation: it is the difference between the population that self-identifies as white *without* any NH self-identification and the population that self-identifies as NH *without* any white self-identification. Despite the unusual interpretation for the primary racial association, the multiracial interpretations are straightforward. For example, β_W is the average difference in SRH between any and no white self-identification. Thus, β_W also corresponds to the average difference in SRH between the white-NH and NH populations.

The significant advantage of the ‘all self-identified races’ approach is the reduced sample size requirements. Yet the approach may also simplify the study design as multiracial participants may self-identify asmultiple races but refuse to self-identify the ‘primary racial representation’ (Boutte-Heinilumoa 2012). The ‘all self-identified races’ approach would allow straightforward inclusion of these participants whereas the ‘primary race’ and ‘secondary race’ approaches would treat data for these participants as missing.

In summary, there is an advantage and disadvantage for each approach. The ‘primary race’ option has the advantage of simplified analyses, but fails to capture variability in multiracial populations. Thus, it would fail to capture health disparities among multiracial populations. The ‘primary and secondary approach’ allows for a consideration of multiracial identity, but

necessitates large sample sizes and/or may lead to missing data. The ‘all self-identified race’ approach alleviates sample size concerns and reduces the potential for missing data at the cost of flexibly modeling multiracial health disparities.

3. Comparing the approaches empirically with the 2012 H-BRFSS

3.1. The 2012 H-BRFSS

The 2012 H-BRFSS is the Hawaii component of the national survey conducted yearly by the Centers for Disease Control and Prevention. The BRFSS collects data on health conditions and behaviors. In contrast to the national BRFSS, the H-BRFSS asks detailed race and ethnicity questions that better captures the highly diverse and multiracial population of Hawaii. This is important as many who live in Hawaii self-identify as specific racial/ethnic groups with historically unique immigration patterns. This means that ‘multiracial’ in Hawaii can encompass several distinct racial and ethnic groups. For example, Hawaii possesses a relatively large Portuguese population from the state’s sugarcane plantation era that remains culturally distinct from the broader ‘white’ population. See the Web Appendix for a more detailed discussion of these issues in Hawaii. For these reasons, through the rest of the paper, we consider ‘multiracial’ as the self-identification of more than one race/ethnicity, racial/ethnic subgroup, or nationality, and use ‘race’ to include the complex, multifaceted aspects of racial/ethnic identity (Jones and Smith 2001; Bratter and Gorman 2011).

For the H-BRFSS, respondents are asked to self-identify their race/ethnicity with multiple responses accepted. If the respondent self-identifies as more than one race/ethnicity, then the respondent is asked to self-identify the group that ‘best represents’ the respondent’s race/ethnicity. This response is defined as the ‘primary race.’

Despite Hawaii’s diverse and multiracial population, some racial groups were combined in the H-BRFSS due to a small number of respondents. In particular, racial groups with less than 100 respondents were collapsed into three different ‘other’ racial groups that maintained geographical and/or cultural cohesion. Non-NH Pacific islanders were grouped into an ‘Other PI’ group, Southeast-Asians and Indians were grouped into an ‘Other Asians’ group, and the last ‘Other’ group represented respondents who self-identified as ‘unspecified other.’ The Web Appendix summarizes the number of respondents in the H-BRFSS for each racial/ethnic group.

We were interested in multiracial health disparities in SRH. SRH is the response to ‘would you say that in general your health is…’ with five potential options on a Likert scale ranging from excellent (1) to poor (5). SRH is commonly investigated in racial health disparities research (Bratter and Gorman 2011; Sentell et al. 2014; Vargas, Sanchez, and Kinlock 2015) and is associated with mortality and morbidity (Idler and Benyamini 1997).

3.2. Statistical methods

Descriptive statistics for the H-BRFSS are presented as means for continuous variables and frequencies for categorical variables. SRH is often dichotomized in health disparities research (Bratter and Gorman 2011; Sentell et al. 2014; Vargas, Sanchez, and Kinlock

2015). However, we are interested comparing the association of primary racial self-identification with and without adjusting for secondary racial self-identification. The odds ratio for a covariate of interest can change by including a covariate associated with the outcome but unrelated to the covariate of interest, which is due to the non-collapsibility of odds ratios (Greenland, Robins, and Pearl 1999). Since we are particularly interested in the effect of racial disparities after controlling for multiracial status, we do not dichotomize SRH and avoid this issue by instead estimating the association between SRH and multiracial populations with linear regression. We consider three models using different approaches for classifying multiracial status: the common 'primary race' approach, the 'all self-identified races' approach, and the 'secondary race' approach. Besides racial self-identification, we adjust for age and gender.

3.3. Assessing model assumptions

The value added from the distinction between the self-identified primary and secondary race options cannot be tested due to the non-nested nature of the corresponding models, which prevents an *F*-test based on the difference in log-likelihoods. Instead, the 'all self-identified races' and 'secondary race' approaches are compared by:

1. The AIC value of the corresponding models.
2. The average difference in the health outcome between the primary and secondary populations for each racial group, which the 'all self-identified races' approach assumes is zero.

The first comparison based on AIC determines the approach with better predictive performance but fails to provide insight on the reasons for choosing a particular model. In contrast, the second comparison provides context-specific insight on reasons the 'all self-identified races' approach may be inappropriate for a given data set.

Model selection criteria, for example, Akaike Information Criterion (AIC), commonly assesses the fit of competing models while accounting for model complexity. A significant advantage of AIC is the asymptotic equivalence to leave-one-out cross-validation, which commonly approximates predicted error. This implies that the approach with the lowest estimated AIC better predicts self-reported health (SRH). We therefore compare the two approaches with the AIC method proposed by Lumley and Scott (Lumley and Scott 2015), which accounts for the complex survey design of the H-BRFSS when estimating AIC and is implemented in the 'survey' package in R (Lumley 2014). However, AIC provides only an indirect answer to the validity of a particular model, and provides little insight into the context-specific reasons for choosing a particular approach.

For context-specific insight into the validity of the 'all self-identified races' approach, we decompose the expected SRH of the 'all self-identified races' approach into the expected SRH of the primary and secondary races. For example, we are particularly interested in multiracial associations for Native Hawaiians (NH), and the expected SRH for *any* NH self-identification is defined as $E\{Y|AR_{NH} = 1\}$, where Y is the random variable for SRH and AR_{NH} is an indicator for *any* NH self-identification. Let PR_{NH} and SR_{NH} be, respectively,

indicators for primary and secondary NH self-identification. As shown in the Appendix, we obtain

$$E\{Y|\mathrm{AR}_{\mathrm{NH}}=1\}=P(\mathrm{PR}_{\mathrm{NH}}=1|\mathrm{AR}_{\mathrm{NH}}=1)\times E\{Y|\mathrm{PR}_{\mathrm{NH}}=1\}+P(\mathrm{SR}_{\mathrm{NH}}=1|\mathrm{AR}_{\mathrm{NH}}=1)\times E\{Y|\mathrm{SR}_{\mathrm{NH}}=1\},$$

(4)

where the average SRH of a given racial group with the 'all self-identified races' approach is a weighted combination of the average SRH of the self-identified primary and secondary populations. Thus, the 'all self-identified races' approach will fail to capture multiracial variability for a given race when the average SRH is different between the primary and secondary populations. The decomposition suggests two important situations that determine the validity of the 'all self-identified races' approach for a given racial group:

1. If the average SRH is the same between the self-identified primary and secondary populations, then the 'all self-identified races' approach is appropriate.
2. If the average SRH is different between the self-identified primary and secondary populations, then the 'all self-identified races' approach is inappropriate. This is particularly important for racial groups with a relatively large secondary population compared to the primary population.

In order to determine the appropriate situation for a given data set, we have to estimate the proportions of, and the average difference in SRH between, the primary and secondary populations for each racial group. A *t*-test or linear regression can estimate the difference in SRH between the self-identified primary and secondary populations for each racial group. However, the primary and secondary populations for a given racial group may have differential socioeconomic status and/or health behaviors. Thus, if controlling for these factors is important, then we can assess the difference in SRH after removing the influence of socioeconomic status and/or health behaviors. See the Web Appendix for a description of this method.

All analyses were completed in R v3.1.2 (R Core Team 2014) with the 'survey' package accounting for the complex survey design of the H-BRFSS (Lumley 2014), which requires specifying the survey weights and stratum. This study was granted exempt status by the University of Hawaii Institutional Review Board.

4. Results

Table 1 presents the descriptive statistics for both the overall study population and the population stratified by multiracial status. Approximately 29% self-identify as more than one race, that is, multiracial. The multiracial population is younger with lower educational attainment and income levels than the non-multiracial population. In addition, the multiracial population is more likely to smoke or drink heavily and less likely to have a personal doctor than the non-multiracial population.

Table 2 presents the racial self-identification of Hawaii. The four most common self-identified primary racial/ethnic groups are White, Japanese, Filipino, and NH. Among the common primary races, the Japanese have the lowest percent of secondary racial self-identification and multiracial status. In contrast, there are several races with higher percentages of secondary racial self-identification than primary racial self-identification: Chinese, Native Americans, Portuguese, Puerto Ricans, and Others. The relative ordering of the 'all self-identified races' proportions is similar to the self-identified primary races although both the Chinese and Portuguese possess a relatively large increase in this classification scheme due to large secondary populations. Native Hawaiians and Portuguese possess high multiracial self-identification within the primary populations.

Table 3 presents the difference in SRH between the self-identified primary and secondary populations for each racial group. The large multiracial population and the significant difference between the primary and secondary white populations suggest that the 'primary race only' approach is unsuitable for identifying health disparities in Hawaii. In combination with Table 2, the results of Table 3 also provide a framework for identifying problematic racial groups for the 'all self-identified races' approach. For example, the difference in SRH between the primary and secondary Mexican populations is significant and relatively large. Table 2 shows that secondary racial self-identification accounts for over 40% of the overall Mexican self-identification, which demonstrates that the large difference in SRH may lead to inappropriate conclusions with the 'all self-identified races' approach regarding the association of Mexican self-identification with SRH. In total, the differences suggest that the 'all self-identified races' approach is likely inappropriate for the H-BRFSS with particular concerns regarding conclusions for whites, Mexicans, other Asians, and Blacks.

Table 4 presents the average difference in SRH between whites and the other racial groups. AIC suggests that the 'secondary race' approach better fits the H-BRFSS data than the 'primary race only' and the 'all self-identified races' approaches, which confirms the expectations based on Table 3. The 'secondary race' approach possesses meaningful differences in comparison to the common primary race analysis and the 'all self-identified races' approach. For example, the 'secondary race' approach attenuates the average difference in SRH from whites for Native Hawaiians, Portuguese, and Puerto Ricans by over 10% compared to the 'primary race' approach. This is potentially due to the relatively large secondary and multiracial populations for each racial/ethnic group. As predicted by the analysis of Table 3, the 'all self-identified races' approach conflates the average difference for Mexicans due to the differences between the primary and secondary populations.

Table 5 demonstrates the existence of health disparities among multiracial populations. In particular, by focusing on the 'secondary race' approach, the populations that self-identify secondarily as Japanese, Puerto Rican, Mexican, and other PI are associated with significantly worse SRH compared to those without the corresponding secondary self-identification. This demonstrates the importance of properly accounting for multiracial status and considering the differences between primary and secondary self-identification.

5. Discussion

The H-BRFSS analysis demonstrates the importance of investigating multiracial health disparities. Health disparities for Native Hawaiians, Portuguese and Puerto Ricans were attenuated by 10% after accounting for multiracial status. Populations that self-identified secondarily as Japanese, Puerto Rican, Mexican, and other PI had significantly poorer SRH compared to populations without the secondary self-identification. This provides new information for considering health disparities in a multiracial population such as in Hawaii. Specifically, our findings indicate that specific racial/ethnic primary and secondary identities matter in health disparities. Population-based health research that excludes multiracial populations, groups all combinations together in one ‘mixed race’ category, or even assume that partial racial/ethnic identity is the same as full racial/ethnic identity may lead to misleading conclusions about the array of health disparities seen in the community and may result in poorly targeted policies.

The distinct approaches for classifying multiracial identity may be particularly important to Native Hawaiians, who were the largest multiracial group in Hawaii with over 80% self-identifying as multiracial. Native Hawaiians have many health disparities compared to other populations in Hawaii and are understudied (Mau et al. 2009). If considering multiracial status can help us better understand the nature of these health disparities, then this is an important area for future research. Of course, quantifying the health among multiracial identities depends on the decisions made by respondents around their cultural identity and the options provided in survey/questionnaire design. For example, the census changed reporting options for multiracial identity between 1990 and 2000, which resulted in a major increase in Native Hawaiians and American Indians (Pew Research Center 2015). These are important topics to consider in future work, particularly how the changes may impact the effect sizes of multiracial identity on health in various populations.

Additionally, as Perez and Charles (2009) note:

> Many Americans have multiple identities that reflect complex ancestral origins, tribal and communal associations, and varied ideological outlooks on race and culture. In general, people do not change their ethnicities as a matter of fashion, but they may emphasize different aspects depending on the circumstances (Perez and Charles 2009). For instance, a person who identifies as Mexican among relatives might identify as Hispanic at work and as American when overseas.

The health implications of such complexity are also important topics for future research and, in particular, may be relevant to Native Hawaiians. For example, previous research has found that Native Hawaiians with a ‘traditional’ mode of acculturation had poorer health (specifically higher diabetes rates) than those of other acculturation types (integrated, assimilated, or marginalized) (Kaholokula et al. 2008). Acculturation mode may impact the choice of racial/ethnic identity chosen on a survey form as well as the health-related experiences of that identity.

The analysis also demonstrates the importance of choosing the appropriate analytical method. The ‘primary and secondary race’ approach might work particularly well for a

multicultural population like Hawaii. For example, the ‘all self-identified races’ approach suggests significant differences for multiracial populations that self-identify as white or Filipino, which may be inappropriate due to the violated assumptions of the approach (e.g. there is a significant difference in SRH for ‘primary’ and ‘secondary’ White populations in Table 3). However, the ‘all self-identified races’ approach can accurately identify multiracial health disparities in other situations. For example, as shown in the Web Appendix, the ‘all self-identified races’ approach was more appropriate for a similar analysis with the 2012 national BRFSS, which does not possess the racial diversity of the HBRFSS. In the 2012 national BRFSS analysis, the ‘all self-identified races approach’ was able to qualitatively reproduce most of the Bratter and Gorman results with a conceptually simple analysis, only one year of data, and less than a fourth of the total participants (Bratter and Gorman 2011), which illustrates the core strengths of the approach. Thus, the appropriate approach to multiracial health disparities depends on the given data set and, therefore, the strengths and weaknesses of both approaches should be considered; see Table 6 for a summary of the strengths and weaknesses for each approach.

As demonstrated by the H-BRFSS analysis, the flexibility of the ‘secondary race’ approach can capture important multiracial variability in health outcomes. As such, at the study design phase, the collection of the self-identified primary race is generally recommended although some multiracial respondents may refuse to self-identify the ‘primary racial representation.’ In this case, the reasons for refusal should be explored to improve subsequent study design and analysis. As it does not require the self-identified primary race, the ‘all self-identified races’ approach could be used in the case of substantial missing data for primary race. In addition, there is considerable interest in the appropriate collection of racial and ethnic status outside of self-identified race (Bilheimer and Klein 2010). For example, the question of ‘how others perceive your racial status’ is associated with SRH in Latinos (Vargas, Sanchez, and Kinlock 2015). However, the role of perceived racial status within the general context of multiracial health disparities and, in particular, self-identified race is understudied and an important area for future research.

The analytical approaches possess regression coefficients with associational interpretations. However, there have been recent developments for extending causal interpretations to racial health disparities (VanderWeele and Robinson 2014). These causal interpretations correspond to the effect of ‘equalizing’ the socioeconomic status of the racial groups. However, the role of multiracial populations within these methods is unclear, and the interpretation may be particularly complicated by different socioeconomic status between multiracial and non-multiracial populations (e.g. see Table 1).

The analytical approaches presented in this paper are specifically designed to be simple and easy-to-implement in practice. These approaches alleviate the issues associated with multiracial disparities research by considering different, but related, definitions of multiracial status. In contrast, more sophisticated statistical methodology may better capture the complex nature of multiracial disparities. For example, the association of secondary races may depend on, or interact with, the primary self-identified race. However, these combinations are extremely difficult to investigate due to small sample sizes. Thus, hierarchical models may better capture the potentially complex relationship of these health

disparities, while also alleviating sample size concerns by shrinking coefficients towards the overall secondary coefficient (Gelman 2006).

In conclusion, we identify multiracial health disparities with two analytical approaches that can potentially overcome the practical and methodological challenges faced by multiracial populations, and we provide practical guidance for determining the appropriate approach for the investigation of multiracial health disparities. The analytical approaches also enable the investigation and identification of multiracial health disparities without prohibitively large sample sizes. Our results additionally illustrate the importance of accounting for multiracial identities in order to fully understand health outcomes. Thus, we provide insights and methods that can allow us to better understand the health needs of the increasingly large multiracial population of the US.

Acknowledgments

Funding

This work was supported by National Institute of General Medical Sciences: [grant number P20GM103466]; National Institute on Minority Health and Health Disparities: [grant number G12MD007601, P20MD000173, U54MD007584].

References

Anderson, NB., Bulatao, RA., Cohen, B. Critical Perspectives on Racial and Ethnic Differences in Health in Late Life. Washington, DC: The National Academies Press; 2004.

Bilheimer LT, Klein RJ. Data and Measurement Issues in the Analysis of Health Disparities. Health Services Research. 2010; 45:1489–1507. [PubMed: 21054368]

Boutte-Heinilumoa, NC. Your Perception, My Reality: The Case of Imposed Identity for Multiracial Individuals. College Station, TX: Texas A&M University; 2012.

Bratter JL, Gorman BK. Does Multiracial Matter? A Study of Racial Disparities in Self-Rated Health. Demography. 2011; 48:127–152. [PubMed: 21347806]

Gelman A. Multilevel (Hierarchical) Modeling: What It Can and Cannot Do. Technometrics. 2006; 48(3):432–435.

Greenland S, Robins JM, Pearl J. Confounding and Collapsibility in Causal Inference. Statistical Science. 1999; 14:29–46.

Griffith DM, Johnson J, Ellis KR, Schulz AJ. Cultural Context and a Critical Approach to Eliminating Health Disparities. Ethnicity and Disparities. 2010; 20(1):71–76.

Humes, KR., Jones, NA., Ramirez, RR. 2010 Census Briefs: Overview of Race and Hispanic Origin. Suitland, Maryland: US Census Bureau; 2011.

Idler EL, Benyamini Y. Self-rated Health and Mortality: A Review of Twenty-Seven Community Studies. Journal of Health and Social Behavior. 1997; 38(1):21–37. [PubMed: 9097506]

Jones, NA., Smith, AS. Census 2000 Brief: The Two or More Races Population. Suitland, Maryland: US Census Burea; 2001.

Kaholokula, Joseph Keawe'aimoku, Nacapoy, Andrea H., Grandinetti, Andrew, Chang, Healani K. Association between Acculturation Modes and Type 2 Diabetes among Native Hawaiians. Diabetes Care. 2008; 31(4):698–700. [PubMed: 18202248]

Kaneshiro B, Geling O, Gellert K, Millar L. The Challenges of Collecting Data on Race and Ethnicity in a Diverse, Multiethnic State. Hawaii Journal of Medicine and Public Health. 2011; 70(8):168–171.

Krogstad, Jens Manuel. Hawaii is Home to the Nation's Largest Share of Multiracial Americans. Washington, DC: Pew Research Center; 2015.

Lumley T. Survey: Analysis of Complex Survey Samples. R package version 3.30. 2014

Lumley T, Scott A. AIC and BIC for Modeling with Complex Survey Data. Journal of Survey Statistics and Methodology. 2015

Mau MK, Sinclair KI, Saito EP, Baumhofer KN, Kaholokula JK. Cardiometabolic Health Disparities in Native Hawaiians and Other Pacific Islanders. Epidemiologic Reviews. 2009; 31:113–129. [PubMed: 19531765]

Mays VM, Cochran SD, Barnes NW. Race, Race-based Discrimination, and health outcomes among African Americans. Annual Review of Psychology. 2007; 58:201–225.

Patten, Eileen. Chapter 1: Estimates of Multiracial Adults and Other Racial and Ethnic Groups Across Various Question Formats. Washington, DC: Pew Research Center; 2015.

Perez, Anthony Daniel, Charles, Hirschman. The Changing Racial and Ethnic Composition of the US Population: Emerging American Identities. Population and Development Review. 2009; 35(1):1–51. [PubMed: 20539823]

Pew Research Center. Multiracial in America: Proud, Diverse and Growing in Numbers. Washington, DC: Pew Research Center; 2015.

R Core Team. R: A Language and Environment for Statistical Computating. Vienna: R Foundation for Statistical Computing; 2014.

Sentell T, Zhang W, Davis J, Baker KK, Braun KL. The Influence of Community and Individual Health Literacy on Self-reported Health Status. Journal of General Internal Medicine. 2014; 29(2): 298–304. [PubMed: 24096723]

United States Department of Health and Human Services. Implementation Guidance on Data Collection Stardards for Race, Ethnicity, Sex, Primary Language, and Disability Status. 2011. aspe.hhs.gov/datacncl/standards/ACA/4302/index.shtml

VanderWeele TJ, Robinson WR. On the Causal Interpretation of Race in Regressions Adjusting for Confounding and Mediating Variables. Epidemiology. 2014; 25(4):473–484. [PubMed: 24887159]

Vargas ED, Sanchez GR, Kinlock BL. The Enhanced Self-reported Health Outcome Observed in Hispanics/Latinos who are Socially-assigned as White is Dependent on Nativity. Journal of Immigrant and Minority Health. 2015; 17(6):1803–1810. [PubMed: 25410381]

Appendix

We decompose of SRH for any self-identification of a given racial group into a weighted combination of the primary and secondary self-identified groups. We assume that there is no missing data and every participant self-identifies a primary race. By iterative expectation, we observe

$$E\{Y|\mathrm{AR}_{\mathrm{NH}}=1\}=E_{\mathrm{PR}_{\mathrm{NH}},\mathrm{SR}_{\mathrm{NH}}|\mathrm{AR}_{\mathrm{NH}}=1}[E\{Y|\mathrm{AR}_{\mathrm{NH}}=1,\mathrm{PR}_{\mathrm{NH}},\mathrm{SR}_{\mathrm{NH}}\}], \quad \text{(A1)}$$

since PR_{NH} and SR_{NH} are binary random variables, we expand equation A1 into

$$\begin{aligned}E\{Y|\mathrm{AR}_{\mathrm{NH}}=1\}=&P(\mathrm{PR}_{\mathrm{NH}}=1,\mathrm{SR}_{\mathrm{NH}}=1|\mathrm{AR}_{\mathrm{NH}}=1)E\{Y|\mathrm{AR}_{\mathrm{NH}}=1,\mathrm{PR}_{\mathrm{NH}}=1,\mathrm{SR}_{\mathrm{NH}}=1\}+\\&P(\mathrm{PR}_{\mathrm{NH}}=1,\mathrm{SR}_{\mathrm{NH}}=0|\mathrm{AR}_{\mathrm{NH}}=1)E\{Y|\mathrm{AR}_{\mathrm{NH}}=1,\mathrm{PR}_{\mathrm{NH}}=1,\mathrm{SR}_{\mathrm{NH}}=0\}+\\&P(\mathrm{PR}_{\mathrm{NH}}=0,\mathrm{SR}_{\mathrm{NH}}=1|\mathrm{AR}_{\mathrm{NH}}=1)E\{Y|\mathrm{AR}_{\mathrm{NH}}=1,\mathrm{PR}_{\mathrm{NH}}=0,\mathrm{SR}_{\mathrm{NH}}=1\}+\\&P(\mathrm{PR}_{\mathrm{NH}}=0,\mathrm{SR}_{\mathrm{NH}}=0|\mathrm{AR}_{\mathrm{NH}}=1)E\{Y|\mathrm{AR}_{\mathrm{NH}}=1,\mathrm{PR}_{\mathrm{NH}}=0,\mathrm{SR}_{\mathrm{NH}}=0\}\end{aligned}$$

(A2)

Note that respondents cannot primarily *and* secondarily self-identify as a given racial group. Thus, PR_{NH} and SR_{NH} cannot overlap each other and, since PR_{NH} and SR_{NH} are also subsets of AR_{NH}, equation A2 further simplifies to

$$E\{Y|\mathrm{AR}_{\mathrm{NH}}=1\}=P(\mathrm{PR}_{\mathrm{NH}}=1|\mathrm{AR}_{\mathrm{NH}}=1)E\{Y|\mathrm{PR}_{\mathrm{NH}}=1\}+P(\mathrm{SR}_{\mathrm{NH}}=1|\mathrm{AR}_{\mathrm{NH}}=1)E\{Y|\mathrm{SR}_{\mathrm{NH}}=1\},$$

which corresponds to equation 3 in the paper and finishes the decomposition.

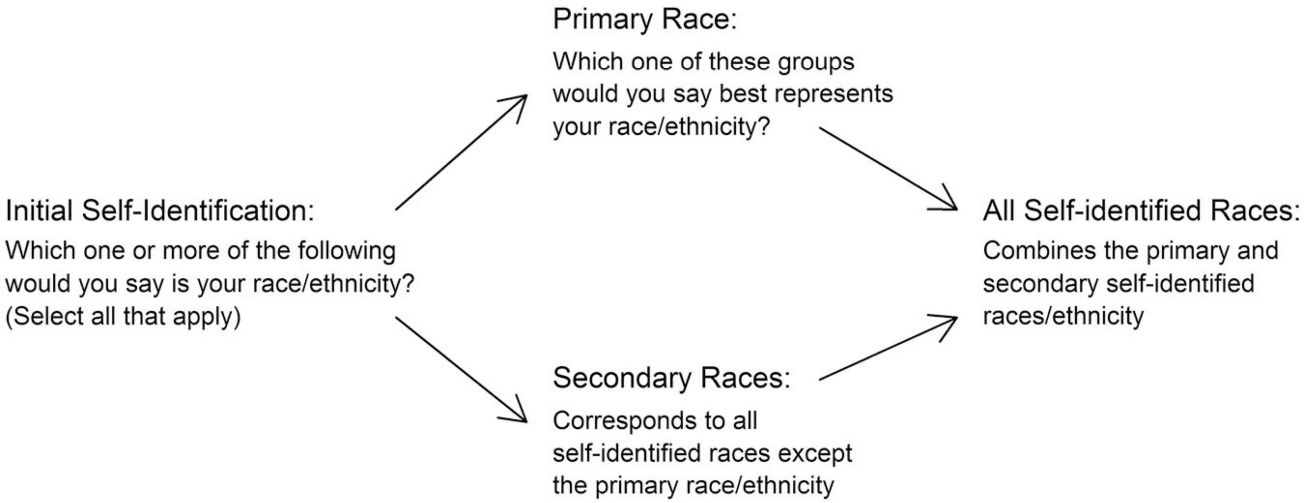

Figure 1.

The flowchart of racial self-identification – the first step asks respondents to self-identify all racial groups. The second step asks participants that self-identified more than one racial group to identify the racial group that 'best represents' the participant's race. The secondary races correspond to the self-identified races other than the primary self-identified race. The final step emphasizes that the 'all self-identified races' approach is combination of primary and secondary self-identified race.

Table 1

The demographics of the 2012 H-BRFSS presented for the overall population and the population stratified by multiracial status.

	Overall	Not multiracial	Multiracial
Variables	**(*n* = 1,079,803)**	**(*n* = 764,108)**	**(*n* = 315,695)**
Multiracial	29.2%	–	–
Age[†]	47.4	49.9	41.4
Male	50.3%	49.5%	50.2%
Exercise in Past 30 Days	81.3%	81.0%	81.9%
Education[a]			
Some high school	9.8%	8.8%	12.3%
High school	30.1%	26.5%	38.8%
Some college	33.3%	33.9%	31.9%
College	26.8%	30.9%	17.0%
Income[†]			
<10,000	5.6%	4.9%	7.2%
10,000–15,000	4.0%	3.3%	5.8%
15,000–20,000	6.4%	5.7%	8.2%
20,000–25,000	7.7%	7.4%	8.4%
25,000–35,000	11.5%	11.4%	11.8%
35,000–50,000	15.6%	15.6%	15.5%
50,000–75,000	15.9%	16.4%	14.8%
>75,000	33.3%	35.4%	28.3%
Personal Doctor[†]	84.7%	85.7%	82.2%
Smoking status[†]			
Current smoker	14.6%	12.4%	19.9%
Former smoker	25.6%	25.9%	24.9%
Never smoked	59.8%	61.8%	55.2%
Heavy drinker[†]	7.4%	6.8%	8.8%

[a] $P < .05$ for differences between multiracial status.

Table 2

The proportion for the self-identified ‘primary’ and ‘secondary’ populations for each racial group.

Race/ethnicity	Primary race[a]	Secondary race[a]	All self-identified races[a]	Multiracial[b]
White	27.9%	11.8%	38.7%	14.7%
Black	1.8%	0.6%	2.3%	20.2%
Chinese	6.8%	9.8%	16.3%	23.9%
Filipino	16.3%	6.2%	21.8%	18.6%
Japanese	22.5%	4.3%	26.0%	14.4%
Korean	1.7%	1.0%	2.6%	15.8%
Native Hawaiian	12.2%	6.5%	18.2%	82.0%
Native American	0.4%	2.3%	2.7%	58.8%
Portuguese	3.2%	5.0%	8.2%	58.9%
Puerto Rican	1.3%	1.5%	2.7%	66.5%
Mexican	1.0%	0.8%	1.8%	25.1%
Other	1.7%	3.2%	4.9%	49.0%
Other Asian	1.1%	0.3%	1.3%	14.7%
Other PI	2.0%	0.8%	2.7%	22.2%

Note: The ‘All Self-Identified Races’ corresponds to any racial self-identification and is not the sum of the primary and secondary races as some multiracial respondents did not self-identify a primary race.

[a]Percentage of the entire population.

[b]Percentage of the given racial group.

Table 3

The difference in general self-reported health between the primary and secondary populations for each racial group.

Race/ethnicity	Difference	95% CI	*P*-value
White	0.35	0.22, 0.48	<.001
Black	0.57	−0.15, 1.29	.123
Chinese	0.04	−0.16, 0.24	.700
Filipino	0.04	−0.22, 0.30	.767
Japanese	0.13	−0.09, 0.36	.239
Korean	−0.13	−0.49, 0.22	.463
Native Hawaiian	−0.11	−0.31, 0.10	.302
Native American	−0.29	−0.92, 0.34	.374
Portuguese	0.04	−0.21, 0.29	.756
Puerto Rican	0.20	−0.25, 0.66	.385
Mexican	1.13	0.58, 1.68	<.001
Other	0.08	−0.31, 0.47	.697
Other Asian	−0.59	−1.32, 0.13	.115
Other PI	0.26	−0.42, 0.93	.457

Notes: Large differences indicate that the ‘All self-identified races’ approach is likely inappropriate for the given racial group, positive differences indicate that the secondary population was associated with worse SRH than the primary population.

Table 4

The difference in SRH between Whites and the given self-identified racial group.

Race/Ethnicity (Ref. = White)	'Primary Race'			'All Self-Identified Races'			'Secondary Race'		
	Estimate	**95% CI**	***P*-value**	**Estimate**	**95% CI**	***P*-value**	**Estimate**	**95% CI**	***P*-value**
AIC		6335			6317			6295	
Black	−0.03	−0.25, 0.19	.789	0.00	−0.23, 0.22	.981	−0.04	−0.26, 0.17	.705
Chinese	0.33	0.17, 0.49	<.001	0.23	0.10, 0.35	<.001	0.32	0.16, 0.48	<.001
Filipino	0.40	0.27, 0.53	<.001	0.33	0.22, 0.45	<.001	0.40	0.27, 0.53	<.001
Japanese	0.24	0.15, 0.33	<.001	0.23	0.15, 0.32	<.001	0.25	0.16, 0.34	<.001
Korean	0.26	0.00, 0.53	.052	0.16	−0.04, 0.35	.116	0.25	−0.03, 0.53	.076
Native Hawaiian	0.51	0.38, 0.64	<.001	0.37	0.22, 0.52	<.001	0.45	0.30, 0.61	<.001
Native American	0.56	−0.03, 1.15	.061	0.29	0.08, 0.50	.007	0.52	−0.04, 1.07	.069
Portuguese	0.33	0.14, 0.51	.001	0.18	0.01, 0.34	.037	0.25	0.05, 0.45	.016
Puerto Rican	0.54	0.18, 0.89	.003	0.50	0.26, 0.73	<.001	0.46	0.15, 0.77	.003
Mexican	0.04	−0.33, 0.40	.835	0.45	0.09, 0.82	.015	0.03	−0.34, 0.40	.873
Other	0.36	0.00, 0.72	.052	0.16	−0.02, 0.34	.080	0.19	−0.05, 0.43	.129
Other Asian	0.83	0.47, 1.20	<.001	0.76	0.42, 1.10	<.001	0.86	0.49, 1.23	<.001
Other PI	0.65	0.26, 1.04	.001	0.63	0.31, 0.94	<.001	0.67	0.28, 1.06	.001

Notes: For the 'secondary races' approach, this corresponds to, for example, aNH in Equation 1, while this corresponds to, for example, $\beta_W - \beta_{NH}$ in Equation 2 for the 'all self-identified races' approach. We are particularly interested in estimates that change across different approaches for a given racial/ethnic group. For example, situations in which the estimate for the 'secondary race' approach is attenuated by 10% compared to the common 'primary race' approach.

Table 5

The association of multiracial identification with for SRH in multiracial populations.

	‘All Self-Identified Races’ approach			‘Secondary Races’ approach		
Race/ethnicity	**Estimate**	**95% CI**	***P*-value**	**Estimate**	**95% CI**	***P*-value**
White	−0.14	−0.23, −0.04	.004	−0.03	−0.16, 0.10	.672
Black	−0.14	−0.37, 0.09	.236	0.28	−0.38, 0.94	.404
Chinese	0.09	−0.02, 0.20	.117	0.01	−0.13, 0.15	.894
Filipino	0.20	0.08, 0.31	.001	0.06	−0.14, 0.26	.551
Japanese	0.10	0.00, 0.20	.047	0.25	0.05, 0.45	.015
Korean	0.02	−0.17, 0.21	.827	−0.11	−0.30, 0.09	.295
Native Hawaiian	0.23	0.12, 0.34	<.001	0.13	−0.03, 0.30	.118
Native American	0.16	−0.02, 0.33	.084	0.06	−0.13, 0.24	.545
Portuguese	0.04	−0.10, 0.18	.569	0.03	−0.16, 0.21	.777
Puerto Rican	0.36	0.14, 0.58	.001	0.42	0.09, 0.75	.012
Mexican	0.32	−0.06, 0.69	.098	0.90	0.43, 1.37	<.001
Other	0.02	−0.15, 0.20	.789	−0.02	−0.22, 0.18	.812
Other Asian	0.62	0.28, 0.97	<.001	0.06	−0.61, 0.73	.861
Other PI	0.49	0.18, 0.80	.002	0.53	0.06, 1.00	.028

Notes: For the ‘secondary races’ approach, this corresponds to, for example, γ_{NH} in Equation 1, while this corresponds to, for example, β_{NH} in Equation 2 for the ‘all self-identified races’ approach. We are particularly interested in multiracial groups that possess significant differences in SRH. For example, secondary Japanese self-identification is associated a significant difference in SRH compared to no secondary Japanese self-identification.

Table 6

A brief summary the advantages and disadvantages of the ‘all self-identified races’ and ‘secondary race’ approaches.

	‘All self-identified races’ approach		‘Secondary race’ approach	
Advantages	**1**	Alleviates sample size concerns	**1**	A more flexible approach
	2	Does not require self-identification of ‘best racial representation’	**2**	Easy interpretation due to straightforward extension of traditional approach
	3	Multiracial associations are easy-to-interpret		
Disadvantages	**1**	Violated assumptions can lead to misleading results	**1**	Potentially small sample sizes in some multiracial combinations
	2	Unusual interpretation of ‘Primary’ associations	**2**	Some participants will refuse to identify the ‘best racial representation’