

Are Comparisons the Answer to Understanding Behavioral Aspects of Aging in Racial and Ethnic Groups?

Keith E. Whitfield,¹ Jason C. Allaire,² Rhonda Belue,³ and Christopher L. Edwards⁴

¹Department of Psychology and Neuroscience, Duke University, Durham, North Carolina.

²Department of Psychology, North Carolina State University, Raleigh.

³Department of Health Policy Administration, The Pennsylvania State University, University Park.

⁴Department of Psychiatry, Duke University Medical Center, Durham, North Carolina.

The knowledge gained from studying diverse populations should help to address inequities and prepare us to deal with the needs of the increasing number of older minorities in this country. At the same time, research that is not properly conducted threatens to lead us astray and misconstrue relationships and outcomes related to behavioral aspects of aging. In this article, we propose that simple comparisons between groups are neither necessary nor sufficient to advance our understanding of ethnic minorities. We discuss common pitfalls conducted in group-differences research, including a specific treatment on the issue of statistical power issues. Our goal is to encourage the use of multiple methodological designs in the study of issues related to racial and ethnic minorities by demonstrating some of the advantages of lesser employed approaches.

Key Words: Ethnicity—Minorities—Racial comparisons—Models of group differences—Statistical power.

THE predicted explosion in the demographic shift toward more ethnic minorities representing a greater proportion of our nation (Angel & Angel, 2006) has made the science of studying race and ethnicity a priority for the scholarly community. The National Institutes of Health (NIH) have made several recent revisions to their guidelines for human subject treatment (NIH, 2001). One of the central points of change in policy is the strong statements and rules about the inclusion of minorities in federally funded research projects. Some scientists believe that this change is for social or political reasons and has no grounding in basic science. Others argue that if we are to adequately and thoroughly test hypotheses and provide answers to questions about America's diverse elders, we must formulate answers based on quality data that reflect the heterogeneity of our population (Curry & Jackson, 2003).

These NIH guidelines are clearly increasing the amount of research that includes ethnic groups other than Caucasians. However, often the results from these studies are not thoroughly discussed relative to race, or only the main effect of race is presented, seldom exploring possible interactions between race and other variables (e.g., health, socioeconomic status, and personality). From these accounts, race is included as a variable to establish or control for between-group differences and not as a central factor of importance. Should every study have as its central question the issue of race? Of course not. When it is important, however, race is frequently conceptualized as Caucasians and others, differences are not well interpreted, and the research is insufficiently powered to detect differences because of small minority samples.

In this article, we argue that comparison studies, however necessary to establish inequities, are not sufficient to advance the science of diversity. Our goal in this article is to facilitate a discussion on how to advance research on psychological

aspects of minority aging by presenting benefits and drawbacks of between-group comparisons and within-group examinations.

Comparison or control groups are necessary to test the effectiveness of interventions and therapeutics; however, when the comparison is between racial groups, the traditional concept of a comparison group, such as a placebo or standard of care control, does not necessarily apply. Often Caucasians are used as the comparison or control group necessary to decipher the importance of the findings from research on an ethnic minority group. Caucasians have traditionally been considered as the "control group" by which an understanding of minorities is gained from observing differences. There are some inherent difficulties with this perspective. First, there is a long history of research that does not include ethnic groups other than Caucasians. The validity of that research is seldom questioned in relation to the generalizability to the population but the validity of the reverse, research focused on a minority groups, is often examined. Second, Caucasians are sometimes thought to be needed in an analysis of ethnic minorities to assess differences. There is an assumption of differences, but different from what? The assumption seems to be that Caucasians represent some sort of standard from which ethnic minorities deviate. Finally, group-difference studies sometimes assume that the same underlying processes produce the outcome of interest. However, the process might be different and therefore leads to a difference in outcomes. For example, African American and Caucasian women have been found to face similar caregiving situations, but African American women report less burden than Caucasian women do (Martin, 2000). These kinds of conceptualizations about racial-differences research have been discussed by Cauce, Coronado, and Watson (1998). They described three models typically used in thinking about and interpreting results from cross-cultural research,

which exemplify the issue of misinterpretation. These models are the (a) Cultural Deviance Model, (b) Cultural Equivalence Model, and (c) Cultural Variant Model.

The Cultural Deviant Model characterizes differences or deviations between groups as deviant and inferior. An example might involve racial-group differences in cognitive aging. An interpretation using this model might suggest, for example, that African Americans do more poorly on cognitive tests because they lack the ability to do the tests. The Cultural Equivalence Model is an improvement over the Cultural Deviance Model in that it proposes that superior socioeconomic status (SES) provides advantages that create superior performance. With the use of the Cultural Equivalence Model, differences in performance on cognitive aging would be described differently. An interpretation of differences in cognitive performance using the Cultural Equivalence Model might suggest that the lack of opportunities to obtain equal education as a result of segregation hampered educational opportunities and achievement, which may account for a large portion of the differences between African Americans and Caucasians on tests of cognitive performance.

The Cultural Deviance Model attributes advantages or superior performance to culture. Putting the onus on culture blames a social group for not having the same ideals, resources, attitudes, and beliefs as the majority culture. Placing culpability on SES shifts the responsibility to social structures that are inherently unbalanced in their distribution of resources. In contrast, the Cultural Variant Model describes differences as adaptations to external forces, exemplifying resilience in the face of oppression. Differences are explained not in relation to a majority or superior group but as culturally rooted internal explanations. The third model by definition allows an appreciation for between-group differences, and it challenges one to explore within-group heterogeneity. Using our example of cognitive aging, an interpretation of the performance differences between African Americans and Caucasians might include a discussion about how culture-fair stimuli were not used, how African Americans may be different because they have a different knowledge base, how among earlier cohorts the expectation was to leave the educational system early to financially support their family, or the fact that aging African Americans tended to live in rural areas where education was more optional than mandatory (Whitfield, 1996; Whitfield & Willis, 1998; Whitfield et al., 1997).

As knowledge about ethnic minorities grows, so has the use of Cultural Variant Models to explain differences found between groups. The Cultural Variant Model is important not only for the design and interpretation of research but also for the translation of research. The presentation of findings in a manner that accurately depicts ethnic minority elders will be more informative for and received better by older minorities. At some level, minority elders know about the phenomena we study and make their own interpretations. It is doubtful that they compare their functioning to that of their aging Caucasian counterparts. Furthermore, given the current and expected growth of ethnic minority groups in the United States such as the predictions for the Hispanic population, the concept of majority-minority comparisons has to be reconsidered because groups who are minorities now will not be in the near future (Angel & Angel, 2006).

Comparisons between two different minority groups may enlighten science and reduce bias by evaluating groups that share similar traits and examining whether the outcomes are different. For example, if we were interested in racial ethnic disparities in the impact of subjugation on subsequent generations' mental health, we might choose to study African Americans and Native Americans. These two groups share several similar features or characteristics, including a loss of familial solidarity, similar educational constraints, and patterns of early mortality.

Currently, most of the research on ethnicity, race, culture, and aging is designed to examine between-group differences in constructs known to be associated with age (Jackson, Antonucci, & Gibson, 1990). Recent areas of research have focused on these distinguishable qualities by addressing the significant differences that are present between ethnic groups. This conceptual or methodological approach has generated a considerable body of literature in the area of racially comparative research on elders. Contemporary researchers contend that these cross-ethnic comparisons have several limitations (Markides, Liang, & Jackson, 1990; Whitfield & Baker, 1999). One limitation in most applications of comparative designed research is that it does not provide insight into the degree of within-group variability. For example, a Hispanic subgroup might include Mexican Americans, Latin Americans, and Puerto Ricans. Each of these cultural subgroups reflects some unique and varying historical culture and levels of assimilation. Inherently, these individual groups are different; by collapsing the groups under one "ethnic umbrella" and then comparing them with Caucasians, important distinctions within each group are lost. These lost distinctions may be very important for interpreting differences across various cultural groups (Whitbourne, Bringle, Yee, Chiriboga, & Whitfield, 2005).

One of the challenges often posed in the study of ethnic groups is the identification of unique or new constructs. John Henryism, for example, is thought to be a behavioral construct that is highly salient, reflecting the personal struggles of African Americans (c.f. James, Hartnett, & Kalsbeek, 1983). Although this unique and interesting behavioral measure has validity for understanding the increased risk for poor health, particularly cardiovascular disease, relative to SES (James, Keenan, Strogatz, Browning, & Garrett, 1992; James, Strogatz, Wing, & Ramsey, 1987), there are relatively few measures that are designed on the basis of a concept that is thought to reflect cultural values and issues more prevalent in African Americans. Identifying new or unique measures or concepts may not be as necessary as grasping how behavioral processes within minority groups make them different or unique in comparisons with Caucasians.

In addition to these limitations, there may be analytic problems to making comparisons. There are three potential problems, issues, or challenges to making comparisons between groups that are often overlooked. The first is differences in sample size. It is not uncommon to observe samples that consist of four to five times as many Caucasians as minorities. The second potential analytic problem involves measurement error (e.g., Ramirez, Ford, Stewart, & Teresi, 2005; Ramirez, Teresi, Holmes, Gurland, & Lantigua, 2006; Teresi & Holmes, 2002). The typical observation that there are mean differences in group

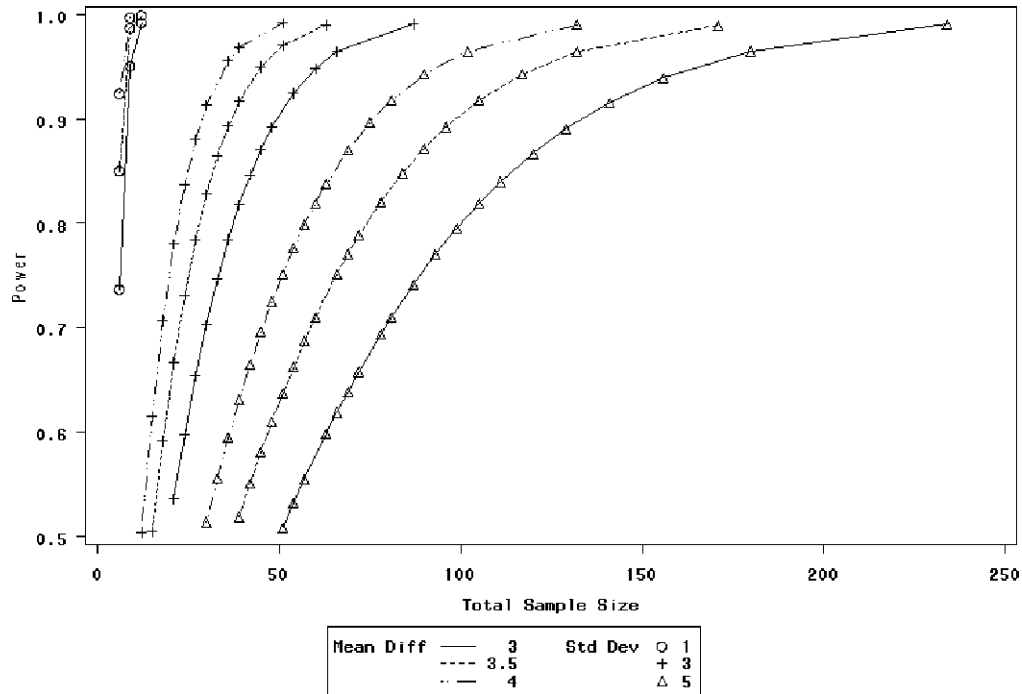


Figure 1. The ratio of African Americans to Caucasians is 1:2.

performance suggests that there may also be differences in measurement error across the groups. This is likely in racial-group comparisons, given the potential differences that arise from dissimilarities in language, history, socialization, and other psychosocial factors.

The third is a basic premise of the analysis of group differences by the use of analysis of variance (ANOVA). One of the assumptions is that there is homogeneity between the two groups. This proposition states that the variance observed in one group must be equal or relative to the group being compared. Studies seldom report tests of homogeneity of variance, and with large sample sizes the violations are usually ignored.

ISSUE 1: SAMPLE SIZE CONSIDERATIONS AND THE POWER TO DETECT DIFFERENCES

The power to detect differences is perhaps one of the most formidable challenges in comparison research. The challenge typically involves getting sufficient numbers of ethnic minorities to participate in research. This has been particularly true for clinical trials research (Whitfield, 2001). Large epidemiological studies and some panel studies have been more successful at getting sufficient numbers of minorities to participate in their studies. What is sometimes lost in the discussions about the sample size of the minority sample is that it typically has to be more than just representative in size to the population from which it is drawn (4% of the population is Pacific Islander so the sample is 4% Pacific Islander). The variances have to be equivalent and the sample size has to be sufficient to detect differences, perhaps even equivalent in size to the comparison or control group. Oversampling can be used to address this issue, but often it is not employed.

To demonstrate the importance of sample size in the statistical power considerations of comparison research, we offer a series of power curves designed to examine the difference in the ratio of African Americans to Caucasians in a hypothetical sample population. The ratio of African Americans to Caucasians varies from figure to figure, with a ratio of 1:2 (one African American for every two Caucasians; twice as many Caucasians as African Americans) in Figure 1, a ratio of 1:4 in Figure 2, and a ratio of 1:6 in Figure 3. As we can see, there is a large change in the statistical power to detect differences between the groups as the ratio for the number of African Americans to Caucasians shrinks. In many studies, good faith efforts to recruit African Americans result in samples that are relative to population estimates. For example, the sample size needed to achieve 90% power with a mean difference between the groups of 4 and a standard deviation of 5 is about 75 if the ratio of African Americans to Caucasians is 1:2. As we can see in Figure 3, the sample size required to achieve the same power doubles when the ratio of African Americans to Caucasians is 1:6, or the common use of the representative sample of African Americans (approximately 15%) to Caucasians.

There are statistical alternatives, such as compound probability, that can be used if multiple samples are available. Compound probability can be used with multiple samples to demonstrate significance if one single sample size is too small (Simon & Burstein, 1985). The goal is to evaluate if there are actual significant differences between samples by examining the probability that differences across multiple assessment are actually significant.

So, the larger the difference in sample size between groups, the larger the required sample size for the minority group to maintain sufficient statistical power. The statistical power to

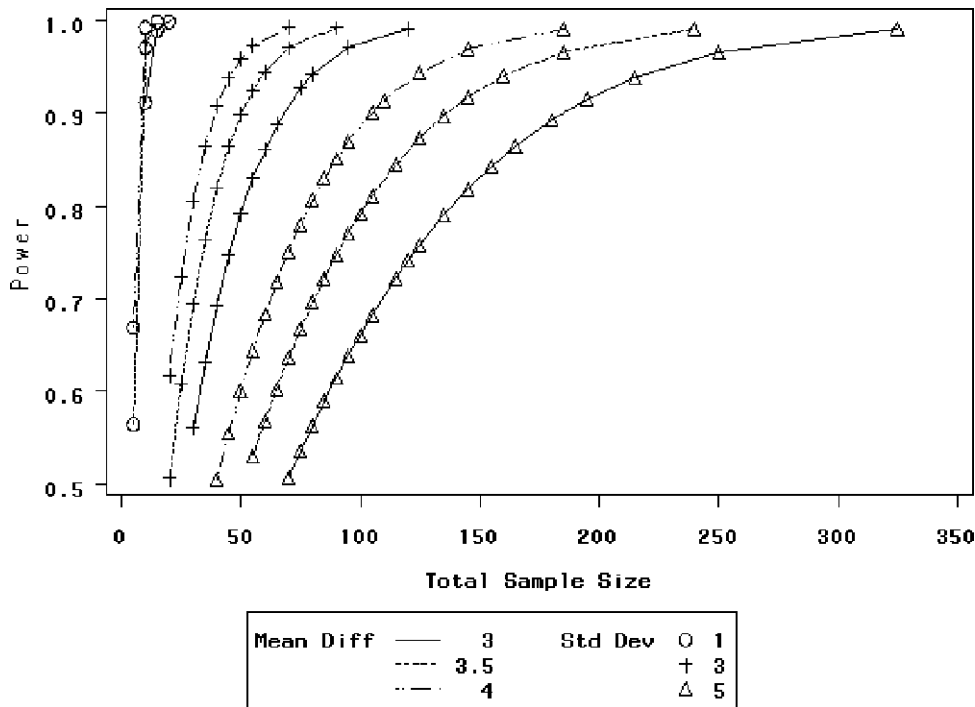


Figure 2. The ratio of African Americans to Caucasians is 1:4.

detect differences is not the only challenge in between-group research. Increasing the number of members of the comparison group increases power, but it does not deal with issues related to measurement error such as precision and variability. Inextricably linked to sample size considerations are issues of measurement and measurement error.

ISSUE 2: MEASUREMENT ISSUES AND ERROR

Precision and variability of measurement are also of concern, particularly when group sample sizes differ. Precision is a measure of how close a parameter estimate is expected to be to the true value of the parameter (Rosenthal, Rosnow, & Rubin, 2000). Precision is then driven by the smaller of the groups. Furthermore, racial and ethnic groups differ in culture, history, experience, living context, beliefs, and cultural norms. These differences can lead to varying interpretations of the same measurement across groups. In order to conduct meaningful between racial-group or ethnic-group comparisons, researchers must ensure that the instrument used to measure the construct of interest has the same meaning across groups, including measurement error and equivalence across groups (e.g., Ramirez et al., 2005). Methods such as item response theory (e.g., Teresi et al., 2007) and variations of factor analysis (e.g., Marshal, Morales, Elliot, Spritzer, & Hays, 2001) are becoming increasingly useful tools for addressing issues in measurement equivalence. Furthermore, item response theory and factor analysis can be employed to evaluate differential item functioning for survey or scale items among racial-ethnic or cultural groups.

It has been shown that racial-ethnic differences in mental health outcomes in diverse groups or elders can sometimes be

explained by differential item functioning (DIF) between racial-ethnic and cultural groups as opposed to true differences in mental health outcomes. Once DIF is controlled, one sees that the differences in these mental health outcomes are not due to race but instead to other covariates such as SES. For example, in a large study of African American and Caucasian elders, researchers found that 89% the difference in cognitive status could be explained by DIF of items on a cognitive measure administered by means of telephone interview to assess cognitive status. The difference in cognitive status was actually explained by differing effects of income on African Americans and Caucasians (Jones, 2003).

DIF can also affect the interpretation of depressive symptoms in elderly persons. DIF has been found to exist for several items on the Center for Epidemiologic Studies-Depression scale. Specifically, measurement and cut-point bias exists between African American and Caucasian elders for items such as "people dislike me." African American elders are more likely to select a higher rating for this item (Yang & Jones, 2007). Overall measurement bias and lack of measurement equivalence as a result of differing racial or ethnic or cultural interpretations of scale items can lead to biased conclusions regarding racial or ethnic and cultural differences in mental health outcomes in diverse groups of elders.

ISSUE 3: ISSUES WITH VARIANCE

Perhaps one of the most straightforward forward problems that can occur in making between-group comparisons is violating assumptions of ANOVA. Comparing means is perhaps one of the most elementary yet common approaches to between-group comparisons (Van de Vijver & Leung, 1997,

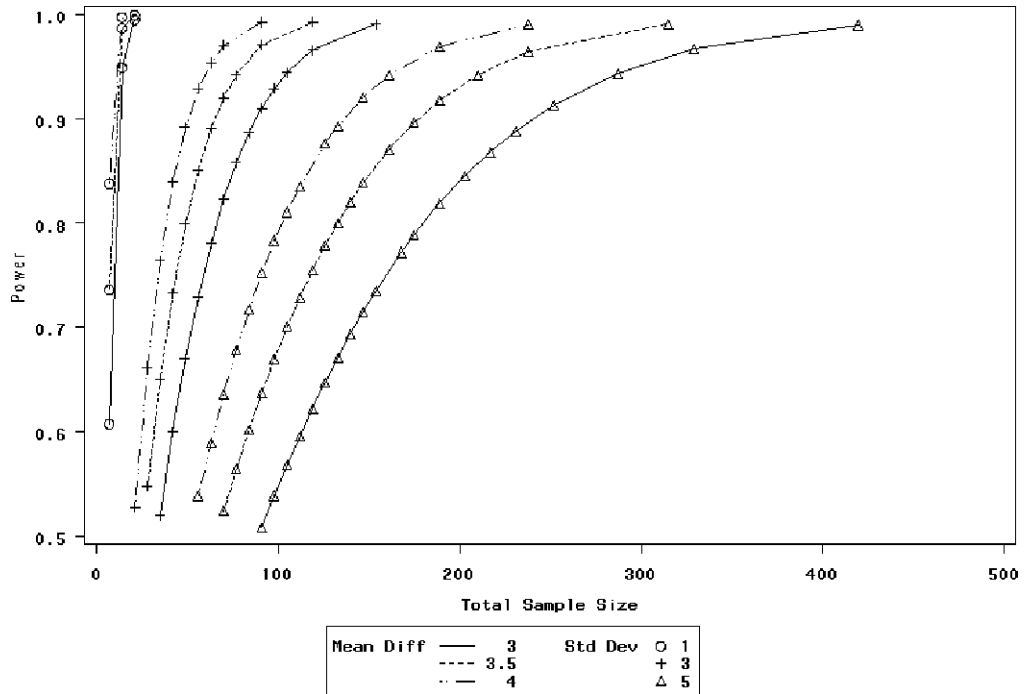


Figure 3. The ratio of African Americans to Caucasians is 1:6.

p. 21). As mentioned earlier, one of the central assumptions of ANOVA is that the groups formed by the independent variable(s) are relatively equal in size and have similar variances on the dependent variable. To demonstrate how this can be evaluated, we drew an example from the initial wave of the American’s Changing Lives data available for public use (House, Lantz, & Herd, 2005). For purposes of this example, we used the cognitive data from the study and plotted it for African Americans and Caucasians. From the scatter plots displayed in Figure 4, we can see that the variances do not look equal. Using the Bartlett test, a commonly used test of homogeneity of variance (cf. Snedecor & Cochran, 1989), we see that the data reveal that our observation that the groups are not homogeneous is correct.

There are a number of data sets that have used oversampling to acquire equal numbers of African Americans and Caucasians and contain measurement of items of interest to researchers interested in social and behavioral aspects of health disparities. For example, the National Center for Health Statistics National Health and Nutrition Interview Survey, which provides an oversample of African Americans and Mexican Americans, and the National Health Interview Survey, and the National Survey on Family Growth, which provides an oversample of African Americans and Hispanics.

We acknowledge that, in the world of research, assumptions of ANOVA are violated more than one might like to see in the literature. Though invariance testing is a prerequisite for group- or time-based comparisons in the context of latent variable

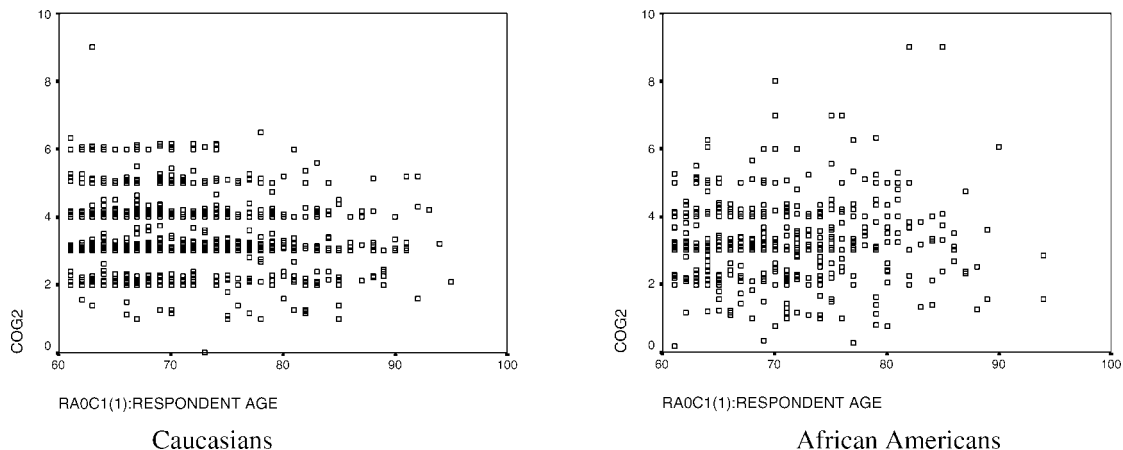


Figure 4. Scatter plots for cognitive data from the America’s Changing Lives study.

analysis, it is almost always a priori assumed and untested within the context of ANOVA. Consequently, statistically significant race effects might reflect measurement differences rather than true differences in the assumed underlying construct. In addition, distributional properties of tests within a specific group may impact between-group mean comparisons. Although statistical tests (ANOVA, *t* tests) are robust against violations to the unequal variances assumption, researchers should be aware of this issue and consider other tests designed to address inequality in variance (e.g., nonparametric or special tests such as the unequal variance *t* test; Ruxton, 2006).

The violation of the assumption of homogeneity in analyzing comparative research is problematic in part because critical covariates are not included in the analysis or interpretations. This adds to the other challenges of doing comparative analyses such as sampling and interpretation of findings. Even though these issues are not too difficult to overcome, attention to these issues provides direct answers to improving comparison research. Within-group analyses also present challenges but avoid some of these common pitfalls.

ARE WITHIN-GROUP DESIGNS THE ANSWER?

The use of within-group designs provides the opportunity to identify the magnitude of heterogeneity within a group and examine how meaningful social variables contribute to this variability. For example, SES is thought to be one of the most important stratifying variables in social science literature (e.g., Adler & Ostrove, 1999). Stratifying by SES may be an important approach to understanding within-group variability by more closely examining how variables such as education and income contribute to social and psychological variables. Techniques such as tests of invariance (e.g., Horn & McArdle, 1992; Meredith, 1993) could be highly useful in shedding light on the variability within ethnic groups and could lead to understanding how variables differ in their contribution to psychological processes. Furthermore, within-group investigations can provide critical information about the processes underlying specific behaviors that is lost in racially comparative analyses and identifying measurement and epistemological differences in the processes under study.

It should be noted that within-group analysis, like between-group analysis, of race differences is not a sufficient singular approach to understanding human behavior. Instead, we consider it the necessary step to describing and understanding behavior, the results of which can more accurately guide any subsequent between-group analysis. For instance, take the finding that African American elders tend to perform worse on measures of cognitive functioning (e.g., Heaton, Ryan, Grant, & Matthews, 1996; Kush et al., 2001) and on cognitive screening tests or batteries (e.g., Manly et al., 1998; Manly, Jacobs, Touradj, Small, & Stern, 2002; Patton et al., 2003; Unverzagt et al., 1996; Whitfield et al., 2000; Zsembik & Peek, 2001). Recent within-group research suggests that age differences in cognition are not associated with the years of education but whether schooling occurred in a segregated versus desegregated school environment (Allaire & Whitfield, 2004). Consequently, a subsequent between-group analysis should determine if school environment is responsible for observed race differences.

Within-group studies can be particularly helpful when researchers are designing interventions and treatments targeted toward racial or ethnic minorities. Understanding heterogeneity within race allows an intervention scientist to develop intervention components that will be effective for the target group. Examination of between-group heterogeneity alone may not provide sufficient information on important characteristics of the target group that would allow prevention scientists to formulate the most effective interventions.

SUMMARY

The initial question that we posed was this: Are comparisons the answer to understanding behavioral aspects of aging in racial and ethnic groups? Our goal in this article was to support the broadening of approaches used in the study of racial and ethnic minority elders by suggesting that comparisons alone are not the answer to understanding aging among minorities. Our intent was to challenge the status quo so that more significant advances in the science of minority aging can be realized. As we have presented here, there are scholarly, justified reasons for using multiple approaches and going beyond simple between-group comparisons. We utilized only some of the excellent examples from the cross-cultural literature to make our points concerning research design and conceptualizations. We also highlighted some of the pitfalls that plague some of the past cross-cultural research of older minorities. It is these issues that limit our ability to fully appreciate aging in different social contexts and groups. The convergence of design and approaches, statistical techniques, and measures should provide greater overlap in results from across studies. For example, studies that include both within-group and between-group approaches are exemplars of the benefits of adequate sampling and statistical and conceptual examination. For instance, researchers have used data from the MacArthur Studies of Successful Aging and have utilized both approaches in analyses of longitudinal data on cognitive aging (between-group analysis, Albert et al., 1995; within-group analysis, Whitfield, 1996). Multiple approaches will also allow for direct comparisons of groups, whether they are collected in one project or across studies, thereby allowing for greater ease of meta-analytic analyses to provide even further insights on aging across race and ethnicity.

We attempted to demonstrate the idea that Caucasians might not be the most logical or necessary contrast group, depending on what question is being asked. To that end, researchers should challenge themselves to ask different questions above and beyond "Are there race differences?" to questions such as "Does education change the course of aging for Hispanics?" Changing the questions and paradigms will help to challenge the current knowledge about ethnicity and aging and significantly advance the science of minority aging.

In this article, we also suggest that between-group studies should not be eliminated but augmented to provide greater insight into the processes that underlie the differences observed between groups. In addition, care has to be taken in blindly making comparisons. More informed theory-based comparisons would benefit the science of minority aging. The models of cultural differences that were discussed earlier provide an intellectual structure to evaluate, design, and interpret results

from between-group comparisons. The creation of a priori hypotheses for why and how differences occur will increase the scientific rigor for minority aging research and research on aging in general. Studying minorities offers unique opportunities not yet fully appreciated by many researchers to understand both basic science and applied science issues in aging. Future research that accurately and appropriately utilizes minorities for understanding processes in aging will advance the field of gerontology.

Science is advanced by evaluating theories in different groups to see if they remain valid and applicable. We have attempted to suggest caution to those who perform such comparisons without a priori criteria for what constitutes a real difference. There are a number of behavioral dimensions of aging such as caregiving, cognition, social support, and other domains that may have qualitative group differences rather than quantitative differences that require careful consideration and alternative approaches such as those we have discussed. Furthermore, sound scientific practice in the study of race or ethnicity should consider how generalizable previous research is and when the criteria for generalizability are not met and therefore require a complete reevaluation (Dilworth-Anderson, Burton, & Turner, 1993).

ACKNOWLEDGMENTS

K.E. Whitfield and J.C. Allaire are supported by a grant from the National Institute on Aging (Grant 1 R01 AG 24108). Special thanks to Jeff Elias, Linda Burton, Roland Thorpe, and the anonymous reviewers for their comments on drafts of this manuscript.

CORRESPONDENCE

Address correspondence to Keith E. Whitfield, Department of Psychology and Neuroscience, 9 Flowers Drive, P.O. Box 90085, Duke University, Durham, NC 27708. E-mail: kwhit1@duke.edu

REFERENCES

- Adler, N. E., & Ostrove, J. M. (1999). Socioeconomic status and health: What we know and what we don't. *Annals of NY Academy of Science*, 896, 3–15.
- Albert, M. S., Jones, K., Savage, C. R., Berkman, L., Seeman, T., Blazer, D., & Rowe, J. (1995). Predictors of cognitive change in older persons: MacArthur Studies of Successful Aging. *Psychology and Aging*, 10, 578–589.
- Allaire, J. C., & Whitfield, K. E. (2004). Relationships among education, age, and cognitive functioning in older African Americans: The impact of desegregation. *Aging, Neuropsychology, and Cognition*, 11(4), 443–449.
- Angel, J. L., & Angel, R. J. (2006). Minority group status and healthful aging: Social structure still matters. *American Journal of Public Health*, 96, 1152–1159.
- Cauce, A. M., Coronado, N., & Watson, J. (1998). Conceptual, methodological and statistical issues in culturally competent research. In M. Hernandez & M. R. Isaacs (Eds.), *Promoting cultural competence in children's mental health services* (pp. 305–331). Baltimore: Brookes.
- Curry, L., & Jackson, J. (2003). Recruitment and retention of diverse ethnic and racial groups in health research: An evolving science. In L. Curry & J. Jackson (Eds.), *The science of inclusion: Recruiting and retaining racial and ethnic elders in health research* (pp. 1–7). Washington, DC: Gerontological Society of America.
- Dilworth-Anderson, P., Burton, L. M., & Turner, W. (1993). The importance of values in the study of culturally diverse families. *Family Relations*, 42, 238–242.
- Heaton, R. K., Ryan, L., Grant, I., & Matthews, C. G. (1996). Demographic influences on neuropsychological test performance. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (2nd ed., pp. 141–163). New York: Oxford University Press.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide in measurement invariance in aging research. *Experimental Aging Research*, 18, 117–144.
- House, J. S., Lantz, P. M., & Herd, P. (2005). Continuity and change in the social stratification of aging and health over the life course: Evidence from a nationally representative longitudinal study from 1986 to 2001/2002 (Americans' Changing Lives Study). *Journal of Gerontology: Psychological Sciences*, 60B, P15–P26.
- Jackson, J. S., Antonucci, T. C., & Gibson, R. C. (1990). Cultural, racial, and ethnic minority influences on aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 103–123). San Diego, CA: Academic Press.
- James, S. A., Hartnett, S. A., & Kalsbeek, W. D. (1983). John Henryism and blood pressure differences in black men. *Journal of Behavioral Medicine*, 6, 259–278.
- James, S. A., Keenan, N. L., Strogatz, D. S., Browning, S. R., & Garrett, J. M. (1992). Socioeconomic status, John Henryism, and blood pressure in Black adults: The Pitt County study. *American Journal of Epidemiology*, 135, 59–67.
- James, S. A., Strogatz, D. S., Wing, S. B., & Ramsey, D. L. (1987). Socioeconomic status, John Henryism, and hypertension in Blacks and Whites. *American Journal of Epidemiology*, 128, 664–673.
- Jones, R. N. (2003). Racial bias in the assessment of cognitive functioning of older adults. *Aging Mental Health*, 7, 83–102.
- Kush, J. C., Watkins, M. W., Ward, T. J., Ward, S. B., Canivez, G. L., & Worrell, F. C. (2001). Construct validity of the WISC-III for White and Black students from the WISC-III standardization sample and for Black students referred for psychological evaluation. *School Psychology Review*, 30, 70–88.
- Manly, J. J., Jacobs, D. M., Sano, M., Bell, K., Merchant, C. A., Small, S. A., et al. (1998). Cognitive test performance among non-demented elderly African Americans and whites. *Neurology*, 50, 1238–1245.
- Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychology Society*, 8, 341–348.
- Markides, K. S., Liang, J., & Jackson, J. S. (1990). Race, ethnicity, and aging: Conceptual and methodological issues. *Handbook of aging and social sciences* (pp. 112–129). New York: Van Nostrand Reinhold.
- Marshall, G. N., Morales, L. S., Elliot, M., Spritzer, K., & Hays, R. D. (2001). Confirmatory factor analysis of the Consumer Assessment of Health Plans Study (CAHPS) 1.0 Core Survey. *Psychological Assessment*, 13, 216–229.
- Martin, C. D. (2000). More than work. *Journal of Family Issues*, 21, 986–1005.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525–543.
- Nesselroade, J. R., & Jones, C. J. (1991). Multi-modal selection effects in the study of adult development: A perspective on multivariate, replicated, single-subject, repeated measures designs. *Experimental Aging Research*, 17, 21–27.
- National Institutes of Health. (2001). NIH policy and guidelines on the inclusion of women and minorities as subjects in clinical research. Available at: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-02-001.html>
- Patton, D. E., Duff, K., Schoenberg, M. R., Mold, J., Scott, J. G., & Adams, R. L. (2003). Performance of cognitively normal African Americans on the RBANS in community dwelling older adults. *Clinical Neuropsychologist*, 17, 515–530.
- Ramirez, M., Ford, M. E., Stewart, A. L., & Teresi, J. A. (2005). Measurement issues in health disparities research. *Health Services Research*, 40, 1640–1657.
- Ramirez, M., Teresi, J. A., Holmes, D., Gurland, B., & Lantigua, R. (2006). Differential item functioning (DIF) and the Mini-Mental State Examination (MMSE): Overview, sample, and issues of translation. *Medical Care*, 44, S95–S106.
- Ruxton, G. D. (2006). The unequal variance *t*-test is an underused alternative to Student's *t*-test and the Mann–Whitney *U* test. *Behavioral Ecology*, 17, 688–690.
- Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. New York: Cambridge University Press.

- Simon, J. L., & Burstein, P. (1985). *Basic research methods in social science* (3rd ed.). New York: Random House.
- Sliwinski, M. J., Hofer, S. M., & Hall, C. (2003). Correlated and coupled cognitive change in older adults with and without preclinical dementia. *Psychology & Aging, 18*, 672–683.
- Snedecor, G. W., & Cochran, W. G. (1989). *Statistical methods* (8th ed.). Ames, IA: Iowa State University Press.
- Teresi, J., & Holmes, D. (2002). Some methodological guidelines for cross-cultural comparisons. In J. H. Skinner, J. A. Teresi, D. Holmes, S. Stahl, & A. L. Stewart (Eds.), *Multicultural measurement in older populations* (pp. 3–11). New York: Springer Press.
- Teresi, J. A., Ocepek-Welikson, K., Kleinman, M., Cook, K. F., Crane, P. K., Gibbons, L. E., et al. (2007). Evaluating measurement equivalence using the item response theory log-likelihood ratio (IRTLR) method to assess differential item functioning (DIF): Applications (with illustrations) to measures of physical functioning ability and general distress. *Quality of Life Research, 16*, 43–68.
- Unverzagt, F. W., Hall, K. S., Torke, A. M., Rediger, J. D., Mercado, N., Gureje, O., et al. (1996). Effects of age, education and gender on CERAD neuropsychological test performance in an African American sample. *Clinical Neuropsychologist, 10*, 180–190.
- Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.
- Whitbourne, S. K., Bringle, J. R., Yee, B. W. K., Chiriboga, D. A., & Whitfield, K. (2005). Ethical research dilemmas with minority elders. In J. E. Trimble & C. B. Fisher (Eds.), *The handbook of ethical research with ethnocultural populations and communities* (pp. 217–240). Thousand Oaks, CA: Sage.
- Whitfield, K. E. (1996). Studying cognition in older African Americans: Some conceptual considerations. *Journal of Aging and Ethnicity, 1*, 35–45.
- Whitfield, K. E. (2001). Commentary: The effects of knowledge, attitudes and significant others on decisions to enroll in a clinical trial on osteoporosis: Implications for recruitment of older African American women. *Journal of the National Medical Association, 93*, 1–3.
- Whitfield, K. E., & Baker, T. T. (1999). Individual differences in aging minorities. *International Journal of Aging and Human Development, 48*, 73–79.
- Whitfield, K. E., Fillenbaum, G., Pieper, C., Seeman, T. E., Albert, M. S., Berkman, L. F., et al. (2000). The effect of race and health related factors on naming and memory: The MacArthur studies of successful aging. *Journal of Aging and Health, 12*, 69–89.
- Whitfield, K. E., Seeman, T. E., Miles, T. P., Albert, M. S., Berkman, L. F., Blazer, D. G., et al. (1997). Health indices as predictors of cognition among older African Americans: MacArthur studies of successful aging. *Ethnicity and Disease, 7*, 127–136.
- Whitfield, K. E., & Willis, S. (1998). Conceptual issues and analytic strategies for studying cognition in older African Americans. *African-American Research Perspectives, 4*, 115–125.
- Yang, F. M., & Jones, R. N. (2007). Center for Epidemiologic Studies–Depression scale (CES-D) item response bias found with Mantel–Haenszel method was successfully replicated using latent variable modeling. *Journal of Clinical Epidemiology, 60*, 1195–200.
- Zsembik, B. A., & Peek, M. K. (2001). Race differences in cognitive functioning among older adults. *Journal of Gerontology: Social Sciences, 56B*, S266–S274.

Received February 4, 2007

Accepted March 26, 2008

Decision Editor: Karen Hooker, PhD