RACIAL AND ETHNIC DIFFERENCES IN THE RELIABILITY AND VALIDITY OF SELF-REPORTED HEALTH STATUS AMONG U.S. ADOLESCENTS

Jason D. Boardman, Ph.D.*

University of Colorado at Boulder

Department of Sociology

Population Program, Institute for Behavioral Science

219 Ketchum Hall

Boulder, CO 80309-0327

303-735-0172 (Phone)

303-492-8878 (Fax)

BOARDMAN@COLORADO.EDU (Email)

* This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (www.cpc.unc.edu/addhealth/contract.html). This paper was made possible by an internal grant from the Population Program of the Institute for Behavioral Sciences at the University of Colorado at Boulder.

Running Head: Self-Reported Health, Measurement

RELIABILITY AND VALIDITY OF SELF-REPORTED HEALTH STATUS AMONG U.S. ADOLESCENTS

ABSTRACT

PURPOSE: This study examines the reliability and validity of standard self-reported health measures among adolescents paying particular attention to differences across racial and ethnic groups. *METHODS:* Using data from Wave 1 of the National Longitudinal Study of Adolescent Health (N = 13,275), this study compares the internal consistency, test-retest reliability, and convergent validity of survey-based health assessments among non-Hispanic white, non-Hispanic black, Asian-American, Hispanic, Native American, and Multi-Racial adolescents. *RESULTS:* Self-reported morbidity demonstrates an acceptable level of internal consistency ($\alpha = .775$) but low test-retest reliability ($\rho=.393$). Likewise, self-rated health demonstrates poor test-retest reliability (K=.272). However, self-rated health and self-reported morbidity operate in predictable ways across measurement occasions providing indirect evidence for the construct validity of these measures. These results are not similar across race/ethnic groups. *CONCLUSIONS:* Adolescent's self-reported health status is temporally and contextually specific and population health estimates should not rely on data obtained from survey questionnaires administered to children and adolescents.

Keywords: Race/Ethnicity, Validity, Reliability, Self-Rated Health, Self-Reported Morbidity, Adolescents.

INTRODUCTION

A large body of research examines the reliability and validity of health assessments obtained from social surveys (1,2). These studies have helped to refine existing survey questionnaires and to clarify important health-related domains, however, this research focuses almost exclusively on adults and little is known about the reliability or validity of survey based health assessments administered to children and adolescents. Likewise, researchers have noted important race/ethnic differentials in the reliability and validity of self-reported health measures (3,4) but to date no existing research has examined these relationships among younger respondents. This paper uses data from the National Longitudinal Study of Adolescent Health to evaluate the internal consistency, test-retest reliability, and construct validity estimates of health questions administered to younger respondents. Because one of the explicit purposes of the Add Health study is to investigate racial and ethnic disparities in health-related outcomes this study purposely oversamples minority respondents. This facilitates a comparison of reliability and validity of survey administered health assessments across a nationally representative sample of black, Hispanic, Asian, Native American, and white adolescents that is not possible with other existing data sets.

OBJECTIVES

The purpose of this study is to address the following four research questions:

1. Is self-rated health a reliable measure of health status among adolescents?
2. Are standard self-reported morbidity items reliable measures of health status among adolescents?
3. Are self-rated health and self-reported morbidity valid measures of health status among adolescents?
4. Does the reliability and validity of self-rated health and self-reported morbidity vary across racial and ethnic groups?

SELF-REPORTED HEALTH STATUS

A number of studies have examined the reliability and validity of self-reported morbidity among adults by comparing an individual's report of their physical well-being to a physician's assessment of their current health status (6-10**).** While the strength of these associations vary according to the specific disease (11), in general, this body of research finds that adult's descriptions of their physical health status and recall of specific health problems closely resembles the reports of trained medical examiners. Indeed, in one study the researchers find self-reported health status to be a stronger predictor of subsequent mortality than physician assessed health status (5) suggesting that self-reported health measures may be appropriate operationalizations of "health" among adults.

Little research has examined the reliability or validity of self-reported health data obtained from children and adolescents. Some have found that survey responses regarding children's health (ages 5-13) demonstrate considerable internal consistency, temporal reliability, and concurrent validity but these analyses used data from the parents of the children and do not include children's self-reports (12). Using the Diagnostic Interview Schedule for Children-Revised to estimate the test-retest reliability of several commonly used operationalizations of well being, some have documented robust kappa coefficients (e.g. > .75) for measures based on parental reports of general anxiety, oppositional defiant behavior, and attention deficit hyperactive disorder among their children (13) but when these tests were administered to children responses were notably less reliable with kappa coefficients ranging from .33 to .54 (13). Reliability was particularly troublesome with respect to temporally specific questions such as duration and onset of the condition. With respect to health related behaviors, there is evidence that adolescents' responses to questions about health promoting activities (e.g., exercise and nutritional awareness) may be more reliable than those regarding health risks (e.g., smoking) (14, 15). This body of work suggests that researchers should interpret results from structured diagnostic assessments from relatively young

children (i.e., elementary school-aged or earlier) and standard survey assessments of health-related behaviors with caution.

To date there remains little empirical evidence regarding the reliability or validity of self-rated health (SRH) and self-reported morbidity (SRM) among adolescents, two of the most common survey based health assessments. These two measures are easily obtained from interviewer administered and self-administered survey questionnaires and are widely believed to be reliable and valid operationalizations of current health status as they consistently predict subsequent health, illness, and mortality among adults (5, 37, 38). The existing work on reliability and validity of health-related measures among adolescents has been restricted to small (e.g., $n < 50$) community-based studies and no existing studies have explicitly investigated either SRH or SRM (39).

In addition, of the studies evaluating these relationships among adults, only a limited number have explored the validity of self-reported health measures across racial groups and those that do emphasize SRH rather than SRM (3, 4). The accurate assessment of physical health status among children is particularly important because adult health status is increasingly conceptualized in terms of a life-course perspective where injuries, physical insults, risky health behaviors, and poor nutrition early in life are believed to affect subsequent health well into adulthood and beyond (16,17). More importantly, researchers continue to document pronounced race differentials in morbidity and mortality despite statistical controls for a number of important social and economic covariates that are believed to mediate these observed differentials (18). In other words, race-specific differentials in physical health status at early ages coupled with chronic stressors throughout the life course (19) may independently contribute to widening race differentials in health among adults. Therefore, the accurate identification of health status among adolescents is particularly important to those concerned with ameliorating persistently adverse health outcomes among racial and ethnic minorities in the US.

METHODS

Data

All data used in these analyses come from wave 1 of the National Longitudinal Study of Adolescent Health (Add Health). Add Health is a school-based, longitudinal study of 83,135 youth in grades 7 through 12 (26). Data for Wave 1 were collected from youth from 80 high schools and 52 middle schools in the years 1994 and 1995 and follow-up in-home surveys were then conducted with 18,924 youth from the original sample (response rate 78.9 percent). The collection of repeated measures among children across contexts (e.g., school and home) make these data particularly useful for the present inquiry. In other words, internal consistency estimates of reliability can be compared to traditional test-retest reliability estimates to more fully describe the ways in which children respond to survey questionnaires about their physical health status. More importantly, the overrepresentation of black, Hispanic, and Asian children enables more refined group specific analyses that are a central concern of this paper. The analyses presented here use only data from students who responded to health questions in both contexts (e.g., school and home). In total, 13,275 youth are used in the analyses. The average age of respondents in these analyses is 15.4 years with a standard deviation of 1.76.

Measures

This study uses two measures of physical health status: (1) *Self-Rated Health (SRH)* and (2) *Self-Reported Morbidity (SRM)*. First, all youth were asked to respond to the question "in general, how would you rate your health?" Response options included the following: (1) "Poor," (2) "Fair," (3) "Good." (4) "Very Good," and (5) "Excellent." The wording and response options for self-rated health were identical in the at-home and in-school questionnaires. Second, self-reported morbidity was obtained from a list of questions regarding specific physical health problems. In-school questionnaires asked the children the following: "Please tell me how often you have had each of the following conditions in the past month:" (1) "a headache", (2) "a stomach ache or an upset stomach", (3) "did you feel really sick", (4) "did you wake up feeling

tired", (5) "did you have skin problems, such as itching or pimples", (6) "were you dizzy", (7) "did you have chest pain", (8) "did you have aches, pains, or soreness in your muscles or joints?" At-home questionnaires asked children to report the same conditions but instead asked "Please tell me how often you have had each of the following conditions in the past 12 months." The measure SRM sums these values across the eight specific measures of morbidity for a minimum possible value of eight and a maximum possible value of 40 reflecting the worst and best health, respectively.

The Add Health study obtains information on the racial identification of respondents from a number of sources including children's self-identification at home, children's self-identification at school, parents' self-identification, parents' identification of their children, and interviewers' identification. In these analyses, children's self-reported racial identification is obtained from the at-home interview. As children were instructed to indicate as many racial groups as they felt appropriate, the Add Health Data set permit more nuanced measures of racial identification than are otherwise available (20). Accordingly, seven racial groups are used in these analyses. Children that listed more than one racial group were coded as "Multi-Racial" (n = 670). In all other cases, children only indicated one racial group. These groups are as follows: (1) Non-Hispanic White (n = 6,909), (2) Non-Hispanic Black (n = 2,601), (3) Asian (n = 944), (4) Hispanic (n = 1,958), (5) Native American (n = 73), and those who reported "other" only (n = 120).

Analytic Strategy

Three standard methods are used to assess the reliability of self-rated health and self-reported morbidity among adolescents. First, a measure of internal consistency (21) is used to gauge *within* context reliability; the extent to which responses to eight physical morbidity questions can adequately capture adolescents' physical health status (as an latent construct) in different settings and across race/ethnic groups. This is an important first step as subsequent analyses will utilize a standardized factor to describe physical morbidity where high scores represent "better health". Second, two different statistical methods are used to evaluate *between*

context reliability which denotes one of the key interests of this paper. As respondents were asked to respond to similar questions in the two interviews (e.g., in school and at home), test-retest and repeated-measures techniques can be applied to gauge the reliability of health assessments administered to adolescents. For SRH, the wording of the question and response alternatives are identical. Subsequently, Kappa statistics (22) are calculated for SRH responses in school and at home for the total sample and for each race/ethnic group separately. Kappa statistics provide a relatively parsimonious gauge of inter-rater and test-retest reliability (23) and describe the extent to which the observed agreement among raters (or similar responses on repeated measures) is greater than what would be expected by chance and is described by the equation 1:

$$K = \frac{P_A - P_E}{1 - P_E} \quad (1)$$

where P_E is the proportion of pairs that are expected to align, by chance, along the main diagonal of a two-way cross tabulation and P_A is the proportion in agreement across the two assessments. The following guidelines are typically used to identify the extent to which the measures of SRH correspond across contexts: ≤ 0 "Poor;" 0 - .2 "Slight;" .2 - .4 "Fair;" .4 - .6 "Moderate;" .6 - .8 "Substantial;" and .8 - 1.0 "Almost perfect." (24)

Response options and the temporal characteristics of the self-reported morbidity measures varied across the in-home and at-school assessments. Respondents were still provided five response options, however, the in-school questionnaire asked children about the "past month" whereas the at-home questionnaire asked children about the "past 12 months." Response options for the in-school interviews were coded accordingly: "Everyday" = 1; "Often" = 2; "Occasionally" = 3; "Rarely" = 4; and "Never" = 5. Likewise, response options for the at-home interviews referenced the past 12 months and were coded accordingly: "Everyday" = 1; "Almost Every Day" = 2; "About Once a Week" = 3; "Just a Few Times" =4; and "Never" =5. Because the wording of questions and the response options are different across the two interviews, the use of Kappa statistics would be inappropriate to assess reliability. Instead, the reliability of SRM is assessed by

calculating intra-class correlation coefficients are total sample and separately for each racial group. All parameter estimates are obtained from a simple multi-level model (equation 2) in which observations (level 1) are nested within adolescents (level 2). Equation 2 is similar to a one-way ANOVA with random effects model where γ_{00} represents the grand mean for a population of scores and error is specified between adolescents (u_{0j}) and between observations (e_{ij}). An intra-class correlation coefficient (ρ) is derived from this model (equation 3) which describes the relative contribution of child-specific (σ_u^2) residual variance to the total residual ($\sigma_u^2 + \sigma_e^2$) variance where observation-level error is captured with σ_e^2. Values for this coefficient range from 0 indicating "no reliability" to 1 indicating "perfect reliability". All models were estimated with SAS 8.2 PROC MIXED (25).

(2) $$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$$

(3) $$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

In addition to reliability this paper also investigates the validity of SRH and SRM. Validity of health measures is often assessed by comparing scores to a "gold standard" such as physician assessment of current health status. However when this standard is not available researchers will often assess the validity of a particular measurement(s) with convergent validity. By examining the extent to which two or more measurements of a particular trait (i.e., self-reported health status) co-vary across repeated measurement occasions researchers can provide indirect evidence that the two measures capture important characteristics of the underlying construct of interest. Multimethod-Multitrait analyses (40) can be used to assess these processes when the construct of interest (e.g., health) can measured two-or more ways with multi-item scales. These measures are then compared against one another and against different constructs that are measured with the same method. Convergent validity is identified when there are positive and significant correlations between measures of the same construct across measurement methods and

the same method across constructs. These methods are particular to scale items and more importantly, they do not specifically address an important aspect of validity that this paper is concerned with. Namely, covariation associated with *changes* across measurement occasions. For example, a large number of adolescents may report significantly lower SRH at home when compared to this same assessment at school. However, the meaning of this difference is properly interpreted in relation to reports of SRM across the two contexts rather than the absolute sense. If there is no corresponding change in the cross-context assessment SRM then these measures can be seen as both unreliable and invalid. If, however, SRM and SRH change in similar ways across contexts, then these measures may take on traditionally defined notions of unreliability but they may also be seen as valid. In this sense, SRM and SRH measures would be capturing an important aspect of health status but the interpretation of the scores can only be understood within each particular context.

(4) $$SRH_{HOME} = b_0 + b_1(SRH_{SCHOOL}) + \sum_{h=2} b_h x_{ih} + e_i$$

$$e_i \sim NID(0, \sigma^2_{e_i})$$

(5) $$SRM_{HOME} = b^*_0 + b^*_1(SRM_{SCHOOL}) + \sum_{h=2} b^*_h x^*_{ih} + e^*_i$$

$$e^*_i \sim NID(0, \sigma^2_{e^*_i})$$

(6) $$\tilde{y}_{e_i} = \tilde{b}_0 + \tilde{b}_1(x_{e^*_i}) + \tilde{e}$$

$$\tilde{e} \sim NID(0, \sigma^2_{\tilde{e}i})$$

In this paper, convergent validity of self-reported health status is assessed with a three-step series of regression models. Equation 4 presents a multivariate ordinary least squares

regression model in which SRH from the in-home observation (SRH_{HOME}) is regressed on SRH from the at-school observation (SRH_{SCHOOL}). This model controls for mother's education and marital status as well as children's age and sex (presented as $\sum_{h=2} b_h x_{ih}$) and specifies error term (e_i) that is normally distributed with a mean of 0 and a variance of $\sigma^2_{e_i}$. In a similar manner, controlling for mother's education and marital status as well as children's age and sex, SRM from the in-home observation (SRM_{HOME}) is then regressed on SRM from the at-school observation and this model specifies and error term (e_i^*) that is normally distributed with a mean of 0 and a variance of $\sigma^2_{e_i^*}$ (see equation 5). Finally, in equation 6, the residuals from the SRM model (e_i^*) are then regressed on the residuals for the SRH model (e_i). If self-reported health measures demonstrate convergent validity, then the slope coefficient ($\tilde{b}_1$) should be a positive and significantly different from 0 ($p<.05$). All models are estimated with the SVYREG procedure in STATA 7.0 which adjusts parameter estimates and subsequent standard errors for the complex sampling design of the Add Health study (27).

RESULTS

Table 1 presents estimates obtained from three statistical methods to evaluate reliability of self-rated health and self-reported morbidity. First, within-context reliability is estimated for both at-school and in-home assessments with a measure of internal consistency (21) and these estimates are presented in the first two columns of Table 1. Among all respondents, the reasonably high value ($\alpha = .775$) indicates that a summary score of standard Likert scale SRM assessments provides a reliable estimate of physical health status among adolescents. Specifically, within each context those who report low levels on some indicators of health are also likely to report low levels on others. Likewise, with the exception of children who listed their race as "other" only, there are no significant race/ethnic differentials in internal consistency of SRM in the at-school or

in-home surveys. It is important to note that SRM appears to be a somewhat less internally consistent measure of health status when the surveys are administered in adolescents' homes (α = .672) instead of their schools. This finding is most likely due to the recall period associated with the questions; the at-school questionnaires asked about health problems in the "past month" whereas the in-home questionnaires asked about problems in the "past year".

[TABLE 1 ABOUT HERE]

The second set of models in Table 1 estimate the reliability of survey-assessed health status among adolescents across social contexts. The third column in Table 1 presents Kappa statistics and illustrates test-retest reliability for self-rated health. According to these results, self-rated health is only a "fairly reliable" (24) measure of health status among adolescents (K = .272). Test-retest reliability is highest among children who reported that their race was "other" only followed by Native American children and non-Hispanic white children and the reliability of SRH appears to be the weakest among multiracial respondents. To more effectively illustrate the observed unreliability of SRH as a measure of adolescents health status Table 2 presents a cross tabulation of SRH responses across contexts. Here, perfect reliability (assuming no change in the underlying construct) would be evident if the responses were all contained along the main diagonal. In other words, we would expect the 137 students who reported their health to be "Poor" at school to also report their health to be "poor" at home. Instead, however, only 12 (9%) of the 137 students that reported "poor" health at school also reported "poor" health at home and 49 of these same 137 students reported their health as either "very good" or "excellent" when asked at home. Likewise, 56 (7%) of the 862 students who listed their health as "fair" at school later reported their health to be "excellent" and another 179 (21%) of these students reported that their health was "very good". Again, these inconsistencies are summarized by the low Kappa coefficient in Table 1.

[TABLE 2 ABOUT HERE]

The final column in Table 1 presents between-context reliability estimates in the form of intra-class correlation coefficients (ρ). These estimates range from 0 to 1 and describe the extent to which variation in SRM is due to variation within children or across contexts where higher levels indicate more between context reliability. As with the results with SRH, SRM does not demonstrate a robust level of reliability; only 39.3 percent of the variation in SRM appears to be due to observations nested within children. And as with SRH these findings suggest that SRM is the particularly unreliable across contexts among non-Hispanic black children. However, the observed unreliability in SRH among Hispanic adolescents does not appear to be as significant when considering SRM. Indeed, following multi-racial adolescents, Hispanic adolescents had the lowest observed test-retest reliability in SRH but nearly the highest test-retest validity in SRM. As with adults (3), the observed unreliability of SRH among Hispanics may have to do with language differences associated with the meaning of health-related questions on survey questionnaires; Hispanic respondents may be more likely to somatize their mental health status and emotional well-being into standard SRM questions when compared to questions about specific physical health conditions (36).

[TABLE 3 ABOUT HERE]

Table 3 presents estimates of convergent validity for self-reported health status. These estimates are obtained from eight regression models in which cross context SRM residuals are regressed on cross context SRH residuals (see discussion above). Overall, respondents' SRM and SRH appear to be valid assessments (b=.149) of overall health status. However, this relationship varies across race/ethnicity. In contrast to test-retest reliability of SRH questions, validity is significantly higher among Hispanic adolescents (b=.208) compared to non-Hispanic white adolescents (b=.133). And whereas reliability estimates obtained the highest observed levels among Native-American adolescents, the estimates presented in Table 2 question the validity of self-reported health measures among these adolescents (b=.059, p<.15). Lastly, as with the school-

based reliability score, the validity estimate is highest for adolescents who reported their race to be "other" only (b=.277, p<.01).

DISCUSSION

Using data from the National Longitudinal Study of Adolescent Health the primary aim of this study was to evaluate the reliability and validity of self-reported health status obtained from questionnaires administered to adolescents. These relationships are tested with two frequently used survey-based health measures (self-rated health and self-reported morbidity) obtained from interviews in two social contexts (at-school and in-home) to evaluate internal consistency, test-retest reliability, and convergent validity. Although self-reported morbidity measures demonstrate acceptable levels of internal consistency, they do not demonstrate acceptable test-retest reliability. As an example, of the 126 adolescents who reported that they had experienced a headache "every day in the past 12 months" in the at-home interview, only 61 (48%) reported earlier that they had a headache every day in the past month when asked at school and 21 (17%) of these adolescents reported that they either "rarely" or "never" had a headache in the past month. The later is also true for self-rated health where only 2238 (56%) of the 4023 adolescents who reported that there health was "excellent" at school reported this same level at home and 64 (<2%) of these adolescents later reported their health to be "fair" or "poor". This is not the case, however, considering the convergent validity of self-reported health status among adolescents. Indeed, according to validity estimates, self-rated health and self-reported morbidity operate in similar and predictable ways across contexts; those adolescents who reported lower self-rated health status at home compared to the in-school questionnaire, also reported lower levels of health when measured by an index tapping physical morbidity. These findings provide an indirect but reassuring evidence of construct validity.

Several important conclusions can be drawn from these findings. First, the test-retest unreliability is important because it suggests that data obtained from survey questionnaires administered to adolescents should not be used as population estimates of adolescent health status.

Self-reported data are often included in general social demographic descriptions of population health status (5, 28) and according to the findings presented here, it is not clear that this information should be obtained from adolescent responses. Second, these findings may help to clarify what is meant by "health" among younger populations. Findings from qualitative research focusing on the meaning that adults attach to standard self-reported health questionnaires suggests that individuals primarily judge their health by reference to specific illnesses, physical functioning, their health-related behaviors, and specific social-psychological resources (30-34). These characteristics tap broad assessments of overall well-being that span both time and context. Among adolescents, however, it appears that context and time factor importantly into their health assessments. In other words, whereas adults may consider a more global concept of health, adolescents may simply be responding to their health status *right here* and *right now*. Likewise, if self-rated health questions tap other important feelings among adolescents (e.g., perceived safety) then context will factor importantly into meaning attributed to the various responses. Accordingly, the observed test-retest unreliability of these measures should not undermine the perceived effectiveness of data such as these to address important social epidemiologic research questions. Rather, researchers simply need to be clear about temporal and contextual specificity when utilizing data such as these to assess the impact of various social processes on adolescents' health status. This may help to explain why estimates of internal consistency demonstrate significantly stronger reliability than test-retest estimates; reliability is evident when context and time are accounted for.

The findings presented in this paper also speak to important issues associated with the social-epidemiology of racial groups. For example, the intra and inter context reliability estimates for adolescents reporting their race/ethnicity to be "other" only are among the highest values obtained. Likewise, one of the highest observed validity estimates was obtained from children who identify with more than racial category. When investigating race/ethnic differentials in physical health, researchers will often delete cases in which they are unable to classify a respondent's

race/ethnicity into standard a priori categories. It is further assumed that difficult to code responses such as "other" only and multiracial identity are assigned randomly and therefore deletion will not bias parameter estimates. If the racial identification of "other" only and multiracial were assigned randomly, then we would expect to see reliability estimates smaller than the overall average. In other words, as race/ethnicity is an important mechanism through which health (35) and the meaning of health (31) operates, then reliability should be higher among individuals in similar groups when compared to individuals who do not represent a group with similar social characteristics and life experiences. This is not the case, however. Indeed, adolescent respondents who list "other" only, at times, appear to have more in common than members of traditionally identified racial groups. As with analyses involving missing data (29), these findings suggest that individuals listing "other" only as their race and those listing more than one race should not be problematized and excluded from the analyses. Rather they should be included in statistical analyses such as these, and more importantly researchers may consider including multiracial respondents and those who report "other" only as distinct racial/ethnic groups.

Finally, there are two important limitations that should be considered when interpreting the results of this study. First, although SRM contains information on a number of physical health conditions, SRM is assessed by response to a single-item. Therefore, with respect to validity, the findings presented in this paper should not be generalized beyond single-item responses. Second, it is important to remember that validity of self-reported health was assessed with a convergent form of construct validity. These methods are most appropriate when "no criterion or universe of content is accepted as entirely adequate to define the quality measured…" (41). In the case of self-reported health status among adolescents, the ideal criterion or "gold standard" is physician assessed health but the Add Health study does not contain health characteristics from trained medical professionals and therefore criterion validity was not assessed. The findings presented in

this paper only pertain to one aspect of validity (construct validity) and should be interpreted accordingly.

REFERENCES

1. Sadana, R. 2000. "Measuring reproductive health: review of community-based approaches to assessing morbidity." *Bull World Health Organ* 78:640-54.

2. Sadana R, Mathers CD, Lopez AD, Murray CJL, Iburg K. 2000. Comparative analyses of more than 50 household surveys on health status. Global Programme on Evidence for Health Policy Discussion Paper Series: No. 15. World Health Organization.

3. Finch, B. K., R. A. Hummer, M. Reindl, and W. A. Vega. 2002. "Validity of self-rated health among Latino(a)s." *Am J Epidemiol* 155:755-9.

4. McGee, D. L., Y. Liao, G. Cao, and R. S. Cooper. 1999. "Self-reported health status and mortality in a multiethnic US cohort." *Am J Epidemiol* 149:41-6.

5. Ferraro, K.F. and M.M. Farmer. 1999. "Utility of Health Data from Social Surveys: Is There a Gold Standard For Measuring Morbidity?" *American Sociological Review* 64:303-315.

6. Altman, R. D. 1991. "Classification of disease: osteoarthritis." *Semin Arthritis Rheum* 20:40-7.

7. Smith, K. W., S. A. McGraw, S. L. Crawford, and J. B. McKinlay. 1993. "Do blacks and whites differ in reporting Rose Questionnaire angina? Results of the Boston Health Care Project." *Ethn Dis* 3:278-89.

8. Harlow, S. D. and M. S. Linet. 1989. "Agreement between questionnaire data and medical records. The evidence for accuracy of recall." *Am J Epidemiol* 129:233- 48.

9. Anderson, D. W., B. S. Schoenberg, and A. F. Haerer. 1988. "Prevalence surveys of neurologic disorders: methodologic implications of the Copiah County Study." *J Clin Epidemiol* 41:339-45.

10. Linet, M. S., S. D. Harlow, J. K. McLaughlin, and L. D. McCaffrey. 1989. "A comparison of interview data and medical records for previous medical conditions and surgery." *J Clin Epidemiol* 42:1207-13

11. Colditz, G. A., P. Martin, M. J. Stampfer, W. C. Willett, L. Sampson, B. Rosner, C. H. Hennekens, and F. E. Speizer. 1986. "Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women." *Am J Epidemiol* 123:894-900.

12.Raat, H., G. J. Bonsel, M. L. Essink-Bot, J. M. Landgraf, and R. J. Gemke. 2002. "Reliability and validity of comprehensive health status measures in children: The Child Health Questionnaire in relation to the Health Utilities Index." *J Clin Epidemiol* 55:67-76.

13. Schwab-Stone, M., T. Fallon, M. Briggs, and B. Crowther. 1994. "Reliability of diagnostic reporting for children aged 6-11 years: a test-retest study of the Diagnostic Interview Schedule for Children- Revised." *Am J Psychiatry* 151:1048-54.

14. Booth, M. L., A. D. Okely, T. Chey, and A. Bauman. 2001. "The reliability and validity of the physical activity questions in the WHO health behaviour in schoolchildren (HBSC) survey: a population study." *Br J Sports Med* 35:263.

15. Abernathy, T. 1997. "Differences in the reported prevalence of adolescents who have never smoked." *Am J Public Health* 87:298-9.

16. Poulton, R., A. Caspi, B. J. Milne, W. M. Thomson, A. Taylor, M. R. Sears, and T. E. Moffitt. 2002. "Association between children's experience of socioeconomic disadvantage and adult health: a life-course study." *Lancet* 360:1640-5.

17. Barker, D. J. 1995. "Fetal origins of coronary heart disease." *Bmj* 311:171-4.

18. Hayward, M.D., E.M. Crimmins, T.P. Miles, and Y. Yang. 2000. "The Significance of Socioeconomic Status in Explaining the Racial Gap in Chronic Health Conditions." *American Sociological Review* 65:910-929.

19. Geronimus, A.T. 1992. "The weathering hypothesis and the health of African-American women and infants: evidence and speculations." *Ethn Dis* 2:207-21.

20. Harris, David R. and Jeremiah Joseph Sim. 2002. "Who Is Multiracial? Assessing the Complexity of Lived Race." *American Sociological Review* 67:614-627.

21. Chronbach, LJ. 1951. “Coefficient Alpha and the Internal Structure of Tests” *Psychometrika,* 16:297-334.

22. Cohen, J. 1968. “Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit” *Psychological Bulletin,* 70:213-20.

23. Hays, RD, Anderson, RT, and Revicki, D. 1998. “Assessing Reliability and Validity of Measurement in Clinical Trials”. Pp. 169-182 in Staqeut, MJ, Hays RD, and Fayers PM eds., *Quality of Life Assessment in Clinical Practice: Methods and Practice*. New York: Oxford University Press.

24. Landis, JR and GG Koch. 1977. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 33:363-74.

25. Littell, R.C., G.A. Milliken, W.W. Stroup, and R.D. Wolfinger. 1996. *SAS System for Mixed Models*. Cary, NC: SAS Institute Inc.

26. Udry, J. R. 1998. The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994-1996 [machine-readable data file and documentation]. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill.

27. Chantala, Kim and Joyce Tabor. 1999. “Strategies to Perform a Design-Based Analysis Using the Add Health Data”. Carolina Population Center, University of North Carolina at Chapel Hill. Available online at: http://www.cpc.unc.edu/projects/addhealth/files/weight1.pdf.

28. NCHS. 2000. *Health, United States, 2000 with Adolescent Health Chartbook*. Hyattsville: National Center for Health Statistics.

29. Gould, J. B., G. Chavez, A. R. Marks, and H. Liu. 2002. "Incomplete birth certificates: a risk marker for infant mortality." *Am J Public Health* 92:79-81.

30. Silverman, M., S. Smola, and D. Musa. 2000. "The Meaning of Health and Not Healthy: Older African Americans and Whites with Chronic Illness." *Journal of Cross-Cultural Gerontology* 15:139-156.

31. Idler, E.L., S.V. Hudson, and H. Leventhal. 1999. "The Meanings of Self-Ratings of Health." *Research on Aging* 21:458-476.

32. Van Doorn, C. 1999. "A Qualitative Approach to Studying Health Optimism, Realism, and Pessmism." *Research on Aging* 21:440-57.

33. Borawski, E. A., J. M. Kinney, and E. Kahana. 1996. "The meaning of older adults' health appraisals: congruence with health status and determinant of mortality." *J Gerontol B Psychol Sci Soc Sci* 51:S157-70.

34. Krause, N.M. and G.M. Jay. 1994. "What Do Global Self-Rated Health Items Measure?" *Medical Care* 32:930-942.

35. Williams, David R. and Chiquita Collins. 1995. "US Socioeconomic and Racial Differences in Health: Patterns and Explanations." *Annual Review of Sociology*:349-386.

36. Angel, R. Guarnaccia PJ. 1989. "Mind, Body, and Culture: Somatization Among Hispanics." *Social Science and Medicine* 28:1229-38.

37. Idler, E. L., L. B. Russell, and D. Davis. 2000. "Survival, functional limitations, and self-rated health in the NHANES I Epidemiologic Follow-up Study, 1992. First National Health and Nutrition Examination Survey." *Am J Epidemiol* 152:874-83.

38. Idler, E. L., S. V. Kasl, and J. H. Lemke. 1990. "Self-evaluated health and mortality among the elderly in New Haven, Connecticut, and Iowa and Washington counties, Iowa, 1982-1986." *Am J Epidemiol* 131:91-103.

39. Schwab-Stone, M., P. Fisher, J. Piacentini, D. Shaffer, M. Davies, and M. Briggs. 1993. "The Diagnostic Interview Schedule for Children-Revised Version (DISC-R): II. Test-retest reliability." *Journal of the American Academy of Child & Adolescent Psychiatry* 32:651-7.

40. Campbell, D.T., and Fiske, D.W. 1959. "Convergent and Discriminant Validity by the multitrait-multimethod matrix. *Psychological Bulletin* 56: 81-105.

41. Chronbach, L.J. and P.E. Meehl. 1955. "Construct Validity in Psychological Tests" *Psychological Bulletin* 52:281-302. Cited in Crocker, L.M. and J. Algina. 1986. *Introduction to classical and modern test theory*. New York: Holt Rinehart and Winston, pp. 218.

TABLE 1. WITHIN CONTEXT AND BETWEEN CONTEXT RELIABILITY OF SELF-REPORTED HEALTH MEASURES AMONG U.S. ADOLESCENTS.

	Within Context		Between Context		Sample Size
Race/Ethnicity	α_{School} [1]	α_{Home} [1]	K_{SRH} [2]	ρ_{SRM} [3]	
Non-Hispanic White	0.774	0.663	0.283	0.335	6909
Non-Hispanic Black	0.761	0.663	0.245	0.317	2601
Hispanic	0.791	0.705	0.237	0.414	1958
Asian	0.75	0.649	0.245	0.374	944
Native American	0.735	0.701	0.395	0.417	73
Multi-Racial	0.775	0.697	0.219	0.422	670
Other Only	0.833	0.654	0.478	0.368	120
Total	0.775	0.672	0.272	0.393	13275

Note: (1) Cell entries represent measures of internal consistency (Cronbach's Alpha's) for Self-Reported Morbidity (SRM) items; (2) Cell entries represent Kappa statistics for inter-rater reliability and test-retest reliability; (3) Cell entries represent intra-class correlation coefficients obtained from a baseline multilevel model where SRM is specified as the dependent variable. All data are weighted to reflect the sample design.

TABLE 2. RELIABILITY OF SELF-RATED HEALTH STATUS: CROSS TABULATION OF SRH ACROSS INTEVIEW CONTEXTS.

	At-School Assessment					
In-Home Assessment	Poor	Fair	Good	Very Good	Excellent	**Total**
Poor	12 (27.3) (8.8)	15 (34.1) (1.7)	11 (25.0) (0.3)	4 (9.4) (0.1)	2 (4.6) (0.1)	**44** **(0.3)**
Fair	35 (4.4) (25.6)	246 (30.9) (28.5)	329 (41.4) (10.0)	122 (15.4) (2.5)	62 (7.8) (1.5)	**794** **(5.9)**
Good	41 (1.3) (29.9)	366 (11.2) (42.5)	1465 (44.8) (44.6)	1011 (30.9) (20.3)	391 (11.9) (9.7)	**3274** **(24.7)**
Very Good	31 (0.6) (22.6)	179 (3.3) (20.8)	1170 (21.8) (35.6)	2652 (49.5) (53.4)	1330 (24.8) (33.1)	**5362** **(40.4)**
Excellent	18 (0.5) (13.1)	56 (1.5) (6.5)	308 (8.1) (9.4)	1181 (31.1) (23.8)	2238 (58.9) (55.6)	**3801** **(28.6)**
Total	**137** **(1.0)**	**862** **(6.5)**	**3283** **(24.7)**	**4970** **(37.4)**	**4023** **(30.3)**	**13275** **(100.0)**

Note: Cell entries represent a cross-tabulation of unweighted frequencies with row percents and column percents, respectively, in parentheses below.

TABLE 3. VALIDITY OF SELF-REPORTED HEALTH STATUS AMONG ADOLESCENTS: CROSS-CONTEXT CORRESPONDENCE OF SELF-REPORTED MORBIDITY AND SELF-RATED HEALTH.

Race/Ethnicity	**b**	**95 % CI**	**R^2**	**Sample Size**
Non-Hispanic White	0.133	(.166, .100)	0.018	6909
Non-Hispanic Black	0.147	(.210, .084)	0.022	2601
Hispanic	0.208	(.302,.114)	0.039	1958
Asian	0.121	(.243,.000)	0.012	944
Native American	0.059	(.247, -.129)	0.003	73
Multi-Racial	0.208	(.322, .094)	0.043	670
Other Only	0.277	(.455, .099)	0.058	120
Total	0.149	(.174, .124)	0.021	13275

Note: Cell entries represent unstandardized OLS slope coefficients and subsequent confidence intervals. Each estimate is obtained from a race-specific regression model in which cross-context SRM residuals are regressed on cross-context SRH residuals. All residual models control for mother's education, mother's marital status, child's age, and child's sex. All data have been weighted and standard errors have been adjusted with replication-based sampling procedures using STATA 7.0.