Depression and Distress in Blacks and Whites in the US:

Testing a Hypothesis to Explain a Double Paradox

David M. Barnes

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
2015

Abstract

Depression and Distress in Blacks and Whites in the US:

Testing a Hypothesis to Explain a Double Paradox

David M. Barnes

This dissertation tested a methodological explanation for a double paradox in psychiatric epidemiology: a lower prevalence of major depression in Blacks than Whites in the US, coupled with equal and higher levels of psychological distress in Blacks. The first paradox is a lower prevalence of major depression in Blacks than Whites. The second paradox is the discordant results from comparing Blacks and Whites on depression and distress. These are paradoxes from the vantage points of, respectively, dominant theory and conceptual and empirical understandings of the relationship between disorder and distress.

The idea that Blacks in the US express depression and distress more somatically than Whites has been in the literature for decades. If true, it could explain the double paradox. A formal diagnosis of major depression requires endorsing a screening symptom, either sad mood or anhedonia, which are both psychological rather than somatic symptoms. To the extent Blacks express depression more somatically than Whites, depression could be disproportionately undercounted in Blacks due to a lower likelihood of Blacks endorsing a screening symptom, adjusting for underlying levels of depression. Measures of distress share symptom content with the diagnostic criteria for depression but typically do not require endorsing screening symptoms. Thus, if Blacks do somatize depression and distress more somatically than Whites, the depression algorithm may produce a greater undercount of depression in

Blacks than Whites, whereas a similar undercount would not occur with distress measures. Accordingly, both paradoxes could be explained.

This dissertation has three main parts. In part one, the double paradox is documented in a systematic literature review. Using data from two nationally representative household samples, parts two and three test whether Blacks express depression and distress, respectively, more somatically than Whites, whether this accounts for a lower likelihood of Blacks endorsing a screening symptom, and if (part two only) this explains the Black-White depression paradox.

The systematic review provides robust evidence of the double paradox. Parts two and three reveal slightly higher levels, respectively, of depression and distress somatization in Blacks than Whites. However, the underlying structure of these small differences provides no evidence of a broad somatization hypothesis in Blacks. Moreover, no evidence is found that the somatization difference inhibits Blacks' endorsement of screening symptoms. One unexpected finding points to subsequent steps to take towards resolving the double paradox.

Table of Contents

List of Figures

Acknowledgements

I want to thank the members of my internal dissertation committee, Professors Sharon Schwartz, Bruce Link, and Lisa Bates, for generously sharing their time, humor, and ideas with me, in the dissertation process and throughout my six years at Columbia. They have been incredibly supportive and congenial mentors, and I feel very lucky that I got to work with them. I also thank Professors Link and Schwartz for their direction of the Psychiatric Epidemiology Training program at Columbia, in which I was a pre-doctoral fellow for five years. Their commitment to the training program, and to psychiatric epidemiology more generally, gave me exceptional training and models to follow.

Chapter 1: Introduction

This dissertation confronts a double paradox in psychiatric epidemiology: non-Hispanic Blacks have a lower prevalence of mental disorder than non-Hispanic Whites (hereafter Blacks and Whites), but equal and higher levels of psychological distress [1–6]. The first paradox is the lower prevalence of psychiatric disorder in Blacks than Whites. The second paradox is discordant Black-White findings when the two groups are compared on mental disorder and distress. These findings, respectively, are paradoxical from the vantage point of social stress theory and from the conceptual and empirical relationships between mental disorder and distress.

Social stress theory is the dominant framework for interpreting relationships between social location and mental health [7–11]. This theory predicts that disadvantaged groups will have worse mental health than more advantaged groups. The social, political, and economic disadvantages Blacks have experienced vis-à-vis Whites in the US, historically and in the present [12–14], make Black-White comparisons a particularly strong test [15] of the social stress paradigm. The finding of better psychiatric health in Blacks than Whites therefore contradicts and potentially undermines this dominant paradigm.

Regarding the second paradox, mental disorder and distress are distinct though overlapping constructs and they are generally positively associated in empirical findings. The finding of less disorder but more distress in Blacks than Whites therefore contradicts these conceptual and empirical underpinnings of the constructs.

This double paradox matters because it is consistently documented in the literature, because Black-White comparisons are a potent test of the social stress paradigm in the US context, and because it casts significant doubt on what we think we know about relationships between social position, mental disorder, and psychological distress.

Various substantive and methodological hypotheses have been proposed to account for the first paradox. Substantive hypotheses start with the premise that the finding of lower disorder in Blacks than Whites is valid and they have tended to posit a protective factor more prevalent in Blacks than Whites. In contrast, methodological explanations presume the finding is *invalid* and that it stems from flawed research. To date, however, these hypotheses have not been sufficiently tested, or test results have not confirmed them, and the paradox persists. The second paradox, by contrast, has received virtually no research attention.

This dissertation pursues a methodological explanation for both paradoxes. It is motivated by the idea that good, or good enough theories are rare and should not be discarded in the face of contravening evidence until all tests of plausible methodological explanations have failed. Contrarily, one could assert that no theory can be expected to predict accurately every time. While this is a defensible position, the Black-White mental disorder findings derive from a critical test of social stress theory and should be confronted head-on rather than tolerated as aberrations. In the vein of theory conservation, this dissertation targets the disorder rather than distress findings (the latter of which, after all, conform to social stress theory predictions) in testing a methodological explanation. Obviously *any* psychiatric diagnosis is too broad a category to explore in one dissertation; therefore, the focus is on major depressive disorder. Major depression is the focal disorder because the Black-White prevalence difference is relatively pronounced for depression, and ironically, distress measures generally draw from

the symptoms of depression more than from other disorders. Given this symptom overlap, and equal

and higher levels of distress in Blacks than Whites, one might therefore expect Blacks to have equal and

higher levels of depression than Whites. That Blacks have a relatively pronounced lower prevalence

than Whites of major depression suggests a disorder where methodological error might be more easily

detected.

The second chapter reports the results from a systematic literature review comparing Blacks and

Whites on major depression and distress from nationally representative US samples. The third and

fourth chapters report results from testing a methodological explanation for the first and second

paradoxes, respectively.

Chapter 2: Documenting a double paradox: a systematic literature review comparing Blacks and Whites in the US on major depression and psychological distress

Introduction

Commentators frequently note that US studies find an equal and lower prevalence of most psychiatric disorders in Blacks than Whites, but also equal and higher levels of psychological distress in Blacks [2–6, 14, 16, 17]. If these sources accurately summarize the published findings, a double paradox emerges. But are they accurate? Because a systematic review of this literature has never been published, the empirical footing of these paradoxes is unclear.

The first paradox is the discordance between the mental disorder findings and what the social stress paradigm leads us to expect. This paradigm is the dominant framework for interpreting relationships between social position and mental health [7–11] and predicts that disadvantaged groups will have worse mental health than more advantaged groups by virtue of greater stressor exposure and access to fewer coping resources. The *social* stress paradigm can be distinguished from the stress paradigm in that the latter considers only the relationship between stressor exposure and health outcomes whereas the former also considers upstream social factors that pattern stressor exposure and coping resources. Blacks' uniquely marginalized political, economic, and social status in the US [12–14] makes Black-White comparisons a particularly strong test [15] of the social stress paradigm. Accordingly, from the vantage point of this paradigm, the consistent finding of an equal and lower prevalence of psychiatric disorder in Blacks than Whites is a clear and non-trivial paradox.

The second paradox is the discordant mental disorder and psychological distress (hereafter distress) findings in Black-White comparisons. Psychiatric disorder and distress are overlapping, though distinct constructs. Both constructs define aversive mental states, are frequently measured with similar symptoms [18, 19], and are phenomenologically related [21, 22]. However, a distinction is often made that disorder represents dysfunction in the individual whereas distress does not assume such internal dysfunction but more often indicates the presence of stressors in the individual's environment to which the expectable response is psychological distress [22, 23].   The constructs can be causally related when, for example, internal mental dysfunctions give rise to psychological distress or chronic distress arising from chronic stressor exposure makes a person more vulnerable to developing internal mental dysfunctions [19–21]. Because of these links, we should expect that those with a psychiatric disorder would score higher on measures of distress than those without a disorder, and that those scoring higher on distress measures would be more likely to have a psychological disorder than those scoring lower. In fact, empirical evidence documents these associations between depression and distress [24–26]. These findings suggest that in between-group comparisons, the group with a higher prevalence of disorder should also have a higher level of distress. The apparent lack of such concordance in Black-White comparisons in the US thus signifies the second paradox.

Because these ostensible paradoxes occur between two groups where the gap in advantages is stark, they cannot be easily dismissed as tolerable exceptions to the apparent rule. That is, because these paradoxes arise in a comparison that clearly exemplifies the elements of the social stress model, they undermine the paradigm and our often tacit [2, 16, 27] reliance on it when interpreting relationships between social location and mental health. It is therefore important to systematically document the published findings comparing Blacks and Whites in the US on mental disorder and distress,

paying special attention to the robustness of the patterns. If there are paradoxes are to be resolved, systematically documenting them is a logical first step.

In this paper, I report findings from a systematic review of the literature estimating the prevalence of major depression and levels of psychological distress in Blacks and Whites in the US. Among disorders, I focus on major depression for four reasons.   First, a cursory literature review suggests that Blacks' lower prevalence than Whites of the multiple psychiatric disorders diagnosed in the large, nationally representative epidemiology studies conducted since 1980 (the advent of the current psychiatric nosology) is particularly marked for major depression. Second, among mental disorders, major depression is especially vulnerable to stressor exposure [28–31], and hence the Black-White depression finding is a particularly strong challenge to the dominant interpretive model. Third, distress measures typically borrow heavily from the diagnostic criteria for major depression; thus, discordance between depression and distress findings in Black-White comparisons is particularly paradoxical. And fourth, because major depression is among the two or three most prevalent psychiatric disorders in the general population [32, 33], paradoxical findings with respect to it have significant implications for public health.

This review draws only on studies using nationally representative samples of adults in the US, the population in which the paradoxes have been primarily noted. Furthermore, although the degree of America's racialized climate may vary geographically, this climate – historically and currently -- is nevertheless ubiquitous. This review also does not consider subgroups defined by immigrant status, ancestry, or any other variable. In a racialized America, physiognomy often trumps these important subgroup differences in shaping life experiences, and therefore crude race comparisons remain telling. Finally, and crucially, results are excluded that adjust for socioeconomic variables such as income,

6

wealth, education, employment, and marital status because these are core explanatory levers, along with inter-personal discrimination, of the social stress paradigm. To include results adjusting for these mediators is to remove key factors that link social location to mental health. Accordingly, only results that adjust at most for sex and age are included.

Methods

The literature review was conducted in PubMed and PsycINFO databases. The search term algorithm (Table 1) was designed to identify articles reporting on Black-White differences in depression or distress in representative samples of the US population. I sequentially culled articles by title, abstract, and full article review, applying the following inclusion criteria: nationally representative US adult samples in which data are reported comparing Blacks and Whites on either major depression or psychological distress. In the full article review phase, articles were retained if they satisfied these inclusion criteria and if the unadjusted results, or adjusted at most for age and sex, were reported. When two or more articles reported results from the same study, articles providing prevalence data were selected over those reporting odds ratios. From on-going or multi-year studies, articles reporting results over a longer range of years were selected over those reporting results from a subset of these years. A secondary literature review was conducted, using these same inclusion criteria, of the references of all articles finally selected from the primary review.

Results were categorized by whether they estimated the occurrence of major depression or distress. "Major depression" is used here as a broad term to encompass both major depressive episode and major depressive disorder. The difference between major depressive episode and major depressive

7

disorder is that the symptoms of a small minority of individuals who meet criteria for a major depressive episode are better accounted for by a psychotic disorder or another mood disorder with manic elements (e.g., manic-depressive disorder), in which case they receive a psychotic or manic-related diagnosis rather than a major depressive disorder diagnosis. Therefore, major depressive episode is not a disorder, per se, though in fact most individuals meeting criteria for a major depressive episode receive a major depressive disorder diagnosis rather than one of the other diagnoses [23]. Because this overlap is large, studies reporting major depressive episode as an outcome were included in this review in addition to those reporting major depressive disorder.

The Patient Health Questionnaire-9 and -8 (PHQ-9 and PHQ-8) are often presented in the literature as measures of major depression [34]. In this review, however, the PHQs were categorized as distress measures because they fall far short of fully implementing the Diagnostic and Statistical Manual of Mental Disorders (DSM) major depression criteria. For example, in the PHQ, symptoms are counted towards the five-symptom minimum required in the DSM major depressive episode algorithm if they were endorsed as occurring during at least half the days during a minimum two-week period, whereas the DSM stipulates that they occur *nearly every day.* Further, the PHQs require that the symptom occurs *at all* during the day whereas the DSM requires that, where relevant, the symptom occurs *most of the day*. As Horwitz and Wakefield [23] persuasively argue, non-disordered distress sometimes satisfies the DSM criteria for major depression, leading to misclassification as the latter; thus, any loosening of the DSM criteria – as occurs with the PHQ -- opens the door still further to false positive diagnoses.

Distress results were subdivided between those comparing Blacks and Whites on the proportions in each group with high distress scores (a threshold determined by each study) and those

comparing the two groups on their mean distress scores. I calculated prevalence ratios and means ratios, respectively, from the results. I also used *openepi.com* to calculate 95 percent confidence intervals around the prevalence ratios when studies provided Black and White sample sizes, and 95 percent confidence intervals around group means when papers provided standard deviations or standard errors but not the confidence intervals themselves. T-tests of differences in mean distress levels were conducted in *openepi.com* when samples sizes and standard deviations or standard errors were provided.

Results

The literature review (schematically summarized in Figure 1) generated 32 articles reporting 44 relevant outcomes. Seven articles [35–41] report 9 Black-White comparative findings on the prevalence of major depression. The remaining 25 articles [4, 42–65] report 35 distress comparisons between Blacks and Whites.

In studies documenting major depression, the Black and White samples ranged in size, respectively, from 666 to 8,245 and from 4,180 to 31,938. These figures for the studies reporting distress outcomes were 198 to 41,056 for Blacks and 1,102 to 333,119 for Whites.

Major depression

Figure 2 and Table 2 summarize the nine major depression comparisons. Blacks have a lower prevalence than Whites in eight comparisons; six of these differences are statistically significant as

indicated by the prevalence ratios and 95 percent confidence intervals (these ranged in width from 0.09 to 0.27 and at the upper end do not extend above 0.91), and in one case a p-value < 0.001. In the one instance in which Blacks have a higher prevalence of major depression than Whites, the 95 percent confidence interval extends well below 1. Regarding diagnostic timeframes, Blacks have a statistically significantly lower prevalence than Whites in all four lifetime comparisons, in two of three past-year comparisons, and there is no meaningful differences between Blacks and Whites in the two past-30 day comparisons.

Distress

Figure 3 and Tables 3a, 3b, and 3c summarize Black-White prevalence ratios of scoring over the cut-points specific to each study. A higher proportion of Blacks than Whites score over the cut-points in 24 of 25 comparisons. Of these 24 comparisons, 19 are statistically significant (based on 95 percent confidence intervals or significance test results (p < 0.05)) and five are not. In the one study reporting a higher prevalence in Whites, the difference is statistically significant.

Figure 4 and Table 4 summarize Black-White distress means ratios. Blacks have higher mean distress scores than Whites in all 10 comparisons. Two of these differences are statistically significant, two are not statistically significant, and six are indeterminate because neither the standard deviations nor standard errors of the mean estimates were provided.

For distress overall, Blacks have higher levels than Whites in 34 of 35 comparisons; of these 34, 21 are statistically significant, seven are not statistically significant, and six are indeterminate. In the one case where Whites have higher distress than Blacks, the difference is statistically significant. The

10

weighted average prevalence ratio was 1.38 and the weighted average means ratio was 1.17. These averages were obtained by weighting each study's Black-White prevalence or means ratio by the study's Black and White sample size, and assigning the average study sample size to studies where sample size is not reported.

In sum, Blacks have a lower prevalence of major depression than Whites in eight of 9 comparisons; six of these differences are statistically significant. In none of the 9 major depression comparisons do Blacks have a statistically significant higher prevalence than Whites. Blacks have higher distress levels than Whites in 34 of 35 comparisons. Of the 29 comparisons in which the statistical significance of the differences can be tested, Blacks are statistically significantly higher than Whites in 21 comparisons, Whites are statistically significantly higher in one comparison, and there is no statistically significant difference in the remaining seven.

Discussion

Blacks have a lower prevalence of major depression than Whites in eight of nine comparisons but higher distress levels in 34 of 35 comparisons. These results from a systematic review of the literature are consistent with the observations in the literature based on cursory reviews [2–4, 6]. In short, psychiatric epidemiology research fairly consistently finds that Blacks have less major depressive disorder but higher distress levels than Whites in the US.

As noted at the outset, this pattern signifies a double paradox. The first paradox, from the perspective of the social stress paradigm, is that Blacks have a lower prevalence of major depression

11

than Whites, despite having a disadvantaged status in the US. The second paradox is that Black-White comparisons of major depression and distress are discordant. Blacks have a lower prevalence of depression than Whites but higher levels of distress, despite evidence in the broader literature of a strong positive association between major depression and distress [24–26].

Both artifactual and substantive explanations have been proposed to resolve the first paradox. Artifactual explanations presume the findings are invalid due to methodological error. Substantive explanations, on the other hand, presume that the lower prevalence of major depression in Blacks than Whites is valid and have tended to posit protective factors thought to be more prevalent in Blacks than Whites, such as religiosity, and high levels of self-esteem, ethnic identity, and social support [4, 36, 66, 67] to account for the pattern. The second paradox, however, undermines these explanations to the extent that the hypotheses fail to account for why factors protective against major depression do not similarly protect against distress. Thus, the second paradox poses challenges to substantive explanations for the first. As well, the social stress paradigm predicts worse mental health outcomes in disadvantaged groups in part *by virtue* of poorer coping resources. To explain the paradox of a lower prevalence of depression in Blacks than Whites by virtue of *better* coping resources simply recreates the paradox at the locus of the hypothesized mediator, and entails a new paradox to explain.

Empirical tests of substantive explanations have not provided support for these hypotheses to date. Examples include examining whether better social support in Blacks than Whites explains Blacks' lower prevalence of major depression [66, 68]. Despite operationalizing social networks in numerous ways, neither set of researchers found support for this explanation. Results from similar tests of self-esteem, ethnic identity, and religiosity as explanatory factors have not been published, to my knowledge. A more recent substantive hypothesis [4] proposes an interaction between race, stress, and poor health

behaviors (e.g., alcohol consumption) such that at higher stressor levels, unhealthy behaviors are more protective against depression in Blacks than in Whites, while simultaneously leading to worse somatic health in Blacks. Tests of this hypothesis have had mixed results [4, 69, 70]. Moreover, a convincing explanation is lacking for why coping behaviors, engaged in to alleviate distress, would protect against psychopathology rather than the distress target [5].

Among methodological explanations for the Black-White depression paradox, one posits that selection bias in the household sample-based studies that document the paradox, disproportionately undercounts depression in Blacks [10, 71–73]. Specifically, the explanation contends that Blacks are disproportionately represented in the groups excluded from household samples (e.g., the incarcerated, homeless, and those living on military bases), which also have a relatively high prevalence of disorder. For this methodological explanation of the Black-White depression paradox to succeed, we would expect to see stronger evidence of the paradox in demographic subgroups where these selection factors are more operant (young males with lower educational achievement), and weaker evidence where they are less operant (older females with higher educational achievement). However, a recent study tested this explanation and found uniform evidence of the paradox across 24 subgroups cross-tabulated by age, sex, and education [74], thus providing evidence inconsistent with this selection bias explanation.

Another methodological explanation for the first paradox suggests that the diagnostic interview for depression used in epidemiologic studies captures depression more effectively in Whites than Blacks [2, 27, 75, 76]. For example, Breslau et al [27] and Uebelacker et al [76] test whether differential item functioning between Blacks and Whites in the diagnostic interview for depression explains any of the paradox. Differential item functioning occurs when two or more groups differ in their probabilities of endorsing specific symptoms, conditioning on the latent construct. Though both detect small levels of

differential item functioning between Blacks and Whites on several symptoms, it is insufficient in both cases to explain a meaningful portion of the depression paradox.

These two tests are hindered, however, by using samples in which all participants have already endorsed at least one of the two screening symptoms for depression (sad mood and anhedonia), and thus selection into the samples is conditioned on a factor many have suggested Blacks and Whites differ on and which is relevant for diagnosis. Specifically, some clinicians and researchers [77–81] have observed that Blacks are more likely than Whites in the US to experience and/or express depression more somatically (i.e., physically), versus more psychologically. A tendency to express depression more somatically would theoretically inhibit endorsing the more psychologically oriented screening symptoms. Because endorsing a screening symptom is required for a depression diagnosis, failing to do so results in skipping out of the remainder of the depression interview. Consequently, those selecting into the full depression interview are more homogenous on a somatization-psychologization continuum than is the full sample that was asked the screen questions. Tests of differential item functioning in this more homogenous sample are therefore theoretically less able to detect differential item functioning between Blacks and Whites

Importantly, if Blacks and Whites do in fact differ on this somatization-psychologization continuum, it could lead to disproportionately undercounting depression in Blacks since diagnosis requires endorsement of one of the two screeners. By contrast, no endorsement of somatic symptoms (which comprise four of the nine diagnostic symptom criteria for major depression) is required for diagnosis. This observation of greater depression somatization in Blacks than Whites in fact suggests a possible explanation for the Black-White depression paradox.

The distress findings reported in this review lend support to this somatization explanation for the first paradox because distress measures typically assign equal weights to all items, while the depression interview gives greater weight to the psychological over the somatic expression of depression. Future studies testing for differential item functioning between Blacks and Whites in the major depression diagnostic interview should sidestep the limitation of the Breslau et al and Uebelacker et al studies [27, 76] by using a design in which the full interview is conducted among the entire sample.

Substantive and methodological explanations for the first paradox have different implications for the social stress paradigm. A core precept of the paradigm is a main effect of social disadvantage on health – that is, disadvantaged groups will have worse health than more advantaged groups -- yet substantive explanations for the first paradox start with the premise that the finding of a lower prevalence of psychopathology in Blacks than Whites is valid. This suggests the conclusion that psychopathology is either positively associated with social advantage or that social advantage is unrelated to psychopathology, both of which are patently inconsistent with the paradigm. If social disadvantage is unrelated to psychopathology, then looking upstream for causes of psychopathology would not need to extend as far as group membership, in which case the social stress paradigm reduces to the stress paradigm.

Accepting the Black-White major depression results (or, more broadly, the psychiatric disorder results) as valid could imply something else, however: that the social stress paradigm is undiminished because psychiatric disorder outcomes imperfectly capture the universe of relevant mental health outcomes [2], or that results contradicting it are tolerable aberrations to the rule [10, 82]. For example, Brown [2] writes: "Racial stratification can cause mental health problems (among both Blacks and Whites) not systematically described in the existing literatures or psychiatric nosology" (p. 293).

Regarding tolerable exceptions, Aneshensel [82] writes: "In sum, location in the social system influences the probability of encountering stressors, which in turn increase the probability of becoming emotionally distressed; these relationships may occur only among some groups, or only under certain conditions" (p. 19). The problem with both arguments is that they risk rendering the social stress paradigm non-falsifiable. To falsify a paradigm or theory is to produce evidence contradicting its predictions, thereby weakening our confidence in it. For a theory to be scientifically useful and not a statement of faith, it must be testable and therefore vulnerable to tests that can shake our confidence in it. To offer caveats that the universe of outcomes is larger than those tested in any given instance, or that a paradigm's predictions may not materialize in all groups, leaves one open to the criticism of non-falsifiability -- to the critique that contravening evidence can always be explained away, thus making the theory invulnerable to falsification. The first caveat would need to reckon with the fact that the Black-White depression paradox extends to most other psychiatric disorders as well, including anxiety and substance use disorders. The second caveat is particularly problematic in Black-White comparisons, which represent an optimal test of the social stress paradigm. Inconsistency between a theory's predictions and the findings may in fact be tolerable in weaker tests of the theory, but when strong tests fail, the theory ought to come under suspicion.

A more conservative approach to resolving the first paradox is to set aside substantive explanations while first testing all plausible artifactual explanations. This approach has the advantage of not disposing of good, or good enough theory prematurely. In any case, results of this systematic review demonstrate that proposed explanations for the first paradox must contend with the second paradox to succeed. To this end, work remains to be done examining differences and commonalities in how psychopathology and distress arise in and are experienced and reported by Blacks and Whites.

Table 1. Search term algorithm

Articles were included for title review if they satisfied criteria A or B, and C, and D, and E.		
A	African	OR
	Black	OR
	Negro	
		AND
	European	OR
	White	OR
	Caucasian	
		OR
B	ethnic	OR
	ethnicity	OR
	race	OR
	"nationally representative"	OR
	"national probability sample"	
		AND
C	"major depression"	OR
	"major depressive disorder"	OR
	"mental disorders"	OR
	"psychiatric disorders"	OR
	"Diagnostic and Statistical Manual of Mental Disorders" [Mesh]	OR
	depression	OR
	"depressive symptomatology"	OR
	"psychological distress"	OR
	distress	OR
	"psychological stress"	
		AND
D	Human [Mesh]	
		AND
E	English	

Table 2. Summary of findings comparing Blacks and Whites on prevalence of major depression.

Source	Study	Data Collection Period	Instrument	Outcome	Black *n*[1]	White *n*[1]	Black Prevalence (95% CI)[2]	White Prevalence (95% CI)[2]	Black-White Prevalence Ratio (95% CI)[3]
Lifetime prevalence									
Weissman et al. (1991)	ECA	1981	DIS	MDD	NR	NR	3.1[4]	5.1	0.61
Blazer et al. (1994)	NCS	1990-1992	CIDI	MDE	931	6,098	11.9 (8.76 - 15.04)	17.9 (16.33 - 19.47)	0.67 (0.56 - 0.80)
Breslau et al. (2006)	NCS-R	2001-2003	CIDI	MDD	717	4,180	10.8 (8.45 - 13.15)	17.9 (16.53 - 19.27)	0.60 (0.48 - 0.75)
Hasin et al. (2005)	NESARC	2001-2002	AUDADIS-IV	MDD	8,245[5]	24,507[5]	8.93 (8.02 - 9.87)	14.58 (14.01 - 15.15)	0.61 (0.57 - 0.66)
Last 12 months prevalence									
Compton et al. (2006)	NLAES	1991-1992	AUDADIS-IV	MDE	5,955[6]	31,938[6]	2.48 (2.01 - 2.95)	3.5 (3.27 - 3.74)	0.71 (0.60 - 0.84)
US DHHS (2001)	NCS	1990-1992	CIDI	MD	666	4,498	8.2 (6.04 - 10.36)	9.9 (8.72 - 11.08)	0.84 (0.64 - 1.09)
Hasin et al. (2005)	NESARC	2001-2002	AUDADIS-IV	MDD	8,245[6]	24,507[6]	4.52 (3.89 - 5.15)	5.53 (5.20 - 1.84)	0.82 (0.73 - 0.91)
Last 30 days prevalence									
Regier et al. (1993)	ECA	1981	DIS	MDE	4,287	12,606[7]	2.5 (1.91 - 3.09)	2.2 (1.81 - 2.59)	1.14 (0.91 - 1.42)
Blazer et al. (1994)	NCS	1990-1992	CIDI	MDE	931	6,098	3.8 (1.84 - 5.76)	4.7 (3.92 - 5.48)	0.80 (0.57 - 1.13)

Notes. US DHHS = United States Department of Health and Human Services; ECA = Epidemiologic Catchment Area; NCS = National Comorbidity Survey; NCS-R = National Comorbidity Survey – Replication; NESARC = National Epidemiologic Survey on Alcohol and Related Conditions; NLAES = National Longitudinal Alcohol Epidemiologic Survey; DIS = Diagnostic Interview Schedule; CIDI = Composite International Diagnostic Interview; AUDADIS-IV = Alcohol Use Disorder and Associated Disabilities Interview Schedule - DSM-IV Version ; MDD = major depressive disorder; MDE = major depressive episode; MD = major depression; NR = not reported.

[1] unweighted

[2] weighted to the US population

[3] prevalence ratios and confidence intervals estimated by DM Barnes

[4] p < 0.001

[5] source: Hasin, DS, et al., 2007

[6] estimated by DM Barnes

[7] non-Black and non-Hispanic (this group is comprised of approximately 94% Whites according to the 1980 US census)

Table 3a. Summary of findings comparing Blacks and Whites on prevalence of high distress, last 12 months

Source	Study	Data Collection Period	Instrument	Outcome	Black n[1]	White n[1]	Black Prevalence (95% CI)[2]	White Prevalence (95% CI)[2]	Black-White Prevalence Ratio (95% CI)[3]	Notes
Last 12 months										
Mellinger et al. (1978)	National drug use survey	1970-1971	HSC (30 distress items)	meeting minimum threshold on at least 2 of 4 dimensions, one of which had to be depression or anxiety	198	2,339	29.4	27.8	1.05 (0.84 - 1.32)	prevalence figures estimated by DM Barnes
Worst month in last 12 months										
Harris et al. (2005)	NSDUH	2001-2003	K6	≥ 13	15,222	94,393	5.32 (4.67 - 5.97)	6.49 (6.26 - 6.73)	0.82 (0.76 - 0.88)	

Notes. NSDUH = National Survey on Drug Use and Health; HSC = Hopkins Symptom Checklist; K6 = Kessler 6; Black = non-Hispanic Black; White = non-Hispanic White; all sample sizes are weighted to the US population; all estimates are unadjusted unless otherwise noted.

[1] unweighted

[2] weighted to the US population

[3] Prevalence ratio confidence intervals estimated by DM Barnes; not estimable when sample sizes not reported

Table 3b. Summary of findings comparing Blacks and Whites on prevalence of high distress, last 30 days

Source	Study	Data Collection Period	Instrument	Outcome	Black n[1]	White n[1]	Black Prevalence (95% CI)[2]	White Prevalence (95% CI)[2]	Black-White Prevalence Ratio (95% CI)[3]	Notes
Last 30 days										
Pratt (2009)	NHIS	1997-2000	K6	≥ 13	13,734	92,270	4.02	3.21	1.25 (1.15 - 1.37)	Black-White OR, adjusting for age and sex = 1.3, 95% CI (1.1 - 1.4)
Dey & Lucas (2006)	NHIS	1998-2003	K6	≥ 13	NR	NR	3.3 (3.01 - 3.59)	2.7 (2.56 - 2.84)	1.22	US-born only
Oraka et al. (2010)	NHIS	2001-2007	K6	≥ 13	21,101	134,265	3.5	2.85	1.23 (1.14 - 1.33)	
Reeves et al. (2011)	NHIS	2009	K6	≥ 13	4,374	16,187	3.8 (3.0 - 4.7)	3.2 (2.8 - 3.6)	1.19 (1.0 - 1.41)	
Strine et al. (2009)	BRFSS	2007	K6	≥ 13	NR	NR	6.1 (5.1 - 7.2)	3.4 (3.2 - 3.5)	1.79	
Reeves et al. (2011)	BRFSS	2009	K6	≥ 13	8,410	68,335	5.4 (4.5 - 6.4)	3.5 (3.2 - 3.9)	1.54 (1.40 - 1.70)	
Dismuke & Egede (2011)	MEPS	2007	K6	≥ 13	2,034	12,673	6.15	5.7	1.08 (0.90 - 1.30)	t-test $p > 0.05$
CDC (2004)	BRFSS	1993-2001	poor mental health days	≥ 14	NR	NR	9.9 (9.6- 10.2)	8.7 (8.6 - 8.8)	1.14	adjusted for age and sex
Chowdhury et al. (2008)	BRFSS	2001-2002	poor mental health days	≥ 14	14,937	153,290	11.2 (10.4 - 12.0)	9.5 (9.3 - 9.8)	1.18 (1.12 - 1.24)	
Mukherjee et al. (2013)	BRFSS	2011	poor mental health days	≥ 14	41,056	396,273	13.46	10.15	1.33 (1.29 - 1.36)	
Zahran et al (2005)	NHANES	2001-2002	poor mental health days	≥ 14	1,009	2,602	4.5 (3.6 - 5.4)	3.6 (3.2 - 4.0)	1.24 (0.87 - 1.75)	
Chowdhury et al. (2008)	BRFSS	2001-2002	days sad, blue, or depressed	≥ 14	7,067	72,107	10.6 (9.3 - 11.9)	7.9 (7.6 - 8.3)	1.34 (1.25 - 1.44)	
Fiscella & Franks (1997)	NHANES	1971-1974	GWB depression subscale	≤13	NR	NR	16.6 (13.9 - 19.29)	8.9 (7.94 - 9.86)	1.86	
Wu & Anthony (2000)	NHSDA	1995-1996	felt sad, blue, or depressed, or lost interest in most things	endorsed ≥ 1	NR	NR	1.11	1.0 (ref)	1.11 (0.79 - 1.56)	odds ratio

Notes. NHIS = National Health Interview Survey; BRFSS = Behavioral Risk Factor Surveillance Survey; MEPS = Medical Expenditure Panel Survey; NHANES = National Health and Nutrition Examination Survey; NHSDA = National Household Survey on Drug Abuse; K6 = Kessler 6; GWB = General Well-Being; Black = non-Hispanic Black; White = non-Hispanic White; NR = not reported; all sample sizes are weighted to the US population; all estimates are unadjusted unless otherwise noted.

[1] unweighted [2] weighted to the US population [3] Prevalence ratio confidence intervals estimated by DM Barnes; not estimable when sample sizes not reported

Table 3c. Summary of findings comparing Blacks and Whites on prevalence of high distress, last 2 weeks and last 7 days

Source	Study	Data Collection Period	Instrument	Outcome	Black n[1]	White n[1]	Black Prevalence (95% CI)	White Prevalence (95% CI)	Black-White Prevalence Ratio (95% CI)[2]	Notes
Last 2 weeks										
Reeves et al. (2011)	NHANES	2005-2008	PHQ9	≥ 10	2,273	4,882	9.7 (7.9, 11.5)	6.2 (5.0, 7.4)	1.57 (1.33 - 1.85)	
CDC (2010)	BRFSS	2006-2008	PHQ8	≥ 5 symptoms more than half the days, including at least one depression screener ("MDD")	17,604	183,563	4.0 (3.6, 4.6)	3.1 (2.9, 3.2)	1.29 (1.20 - 1.39)	
CDC (2010)	BRFSS	2006-2008	PHQ8	2-4 symptoms more than half the days ("Other depression")	17,604	183,563	8.7 (7.9, 9.7)	4.8 (4.6, 5.0)	1.81 (1.72 - 1.91)	
CDC (2010)	BRFSS	2006-2008	PHQ8	any depression (either "MDD" or "other depression")	17,604	183,563	12.8 (11.8, 13.8)	7.9 (7.6, 8.1)	1.62 (1.55 - 1.70)	
Reeves et al. (2011)	BRFSS	2006	PHQ8	≥ 10	15,819	153,642	11.0 (10.1, 12.1)	8.0 (7.7, 8.3)	1.38 (1.31 - 1.44)	
Reeves et al. (2011)	BRFSS	2008	PHQ8	≥ 10	4,837	68,695	12.7 (11.1, 14.6)	7.5 (7.1, 7.9)	1.69 (1.57 - 1.83)	
Last 7 days										
Eaton & Kessler (1981)	NHANES	1975	CES-D	≥ 16	242	2,625	28.5	15.3	1.86 (1.50 - 2.32)	
Jackson et al. (2010)	ACL	1986	CES-D (11 items)[3]	≥ 16	871	1,900	35	21.4	1.64	*t*-test *p* < 0.05;
Jones-Webb & Snowden (1993)	NAS	1984	CES-D	≥ 16	1,947	1,777	20.3	15.1	1.35 (1.17 - 1.55)	Chi-square test p < 0.001

Notes. NHANES = National Health and Nutrition Examination Survey; BRFSS = Behavioral Risk Factor Surveillance Survey; NHANES = National Health and Nutrition Examination Survey; ACL = Americans Changing Lives Survey; NAS = National Alcohol Survey; PHQ = Patient Health Questionnaire; CES-D = Center for Epidemiologic Studies - Depression; Black = non-Hispanic Black; White = non-Hispanic White; NR = not reported; all sample sizes are weighted to the US population; all estimates are unadjusted unless otherwise noted.

[1] weighted

[2] prevalence ratio confidence intervals estimated by DM Barnes

[3] source authors multiplied each endorsed item by 1.818 to produce scores comparable with the conventional 20-item CES-D.

Table 4. Summary of findings comparing Blacks and Whites on distress means

Source	Study	Data Collection Period	Instrument	Black n[1]	White n[1]	Black Mean (95% CI)	White Mean (95% CI)	Black-White Means Ratio	Notes
Last 30 days									
Bratter & Eschbach (2005)	NHIS	1997-2001	K6 (range 0-24)	22,128	107,420	2.41	2.36	1.02	*t*-test *p* > 0.05; regression parameter adjusted for age and sex *p* > 0.05
Kiviniemi et al. (2011)	HINTS	2007	K6 (range 0-24)	580	4,588	1.92 (1.82 - 2.01)	1.82 (1.79 - 1.85)	1.06	SD & SE not reported
Roxburgh (2009)	NHIS	2003	K6 (range 6-30)	3,751	21,247	8.66 (8.54 - 8.78)	8.48 (8.42 - 8.54)	1.07	DM Barnes estimated SE from SD; *t*-test = -2.51, *p* = 0.01
Reeves et al. (2011)	BRFSS	2009	poor mental health days	33,741	333,119	4.1 (3.9 - 4.4)	3.3 (3.2 - 3.4)	1.24	SD & SE not reported
Nuru-Jeter et al. (2008)	MHS	1994	five-item Mental Health Inventory (range 5-30)	995	1,102	11 (10.73 - 11.27)	10.7 (10.45 - 10.96)	1.03	*t*-test = 1.57, *p* = 0.12; D Barnes estimated SE from SD; weighted estimates
Fleischman (2007)	MEPS	2004	K6 + PHQ9 (2 items) + Short Form 2 (2 items)	2,001	12,915	17.62	16.68	1.06	sample sizes estimated by DM Barnes; SD & SE not reported
Last 7 days									
Eaton & Kessler (1981)	NHANES	1975	CES-D (8 items)	242	2,625	10.9	8.4	1.3	SD & SE not reported
Mulia et al. (2008)	NAS	2005	CES-D (8 items)	1,054	3,967	3.96	3.27	1.21	SD & SE not reported
Various time frames in last 30 days									
Kessler & Neighbors (1986)	various	1967 - 1976	various depression screening scales	1,411	8,307	1.24	1	1.24	SD & SE not reported
Kessler & Neighbors (1986)	various	1957 - 1976	various somatization screening scales	3,040	19,404	1.1	1	1.1	SD & SE not reported

Notes. NHIS = National Health Interview Survey; HINTS = Health Information National Trends Survey; BRFSS = Behavioral Risk Factor Surveillance Survey; MHS = Minority Health Survey; MEPS = Medical Expenditure Panel Survey; NHANES = National Health and Nutrition Examination Survey; NAS = National Alcohol Survey; K6 = Kessler 6; PHQ = Patient Health Questionnaire; CES-D = Center for Epidemiologic Studies - Depression; Black = non-Hispanic Black; White = non-Hispanic White; SD = standard deviation; SE = standard error; all sample sizes are weighted to the US population; all estimates are unadjusted unless otherwise noted.

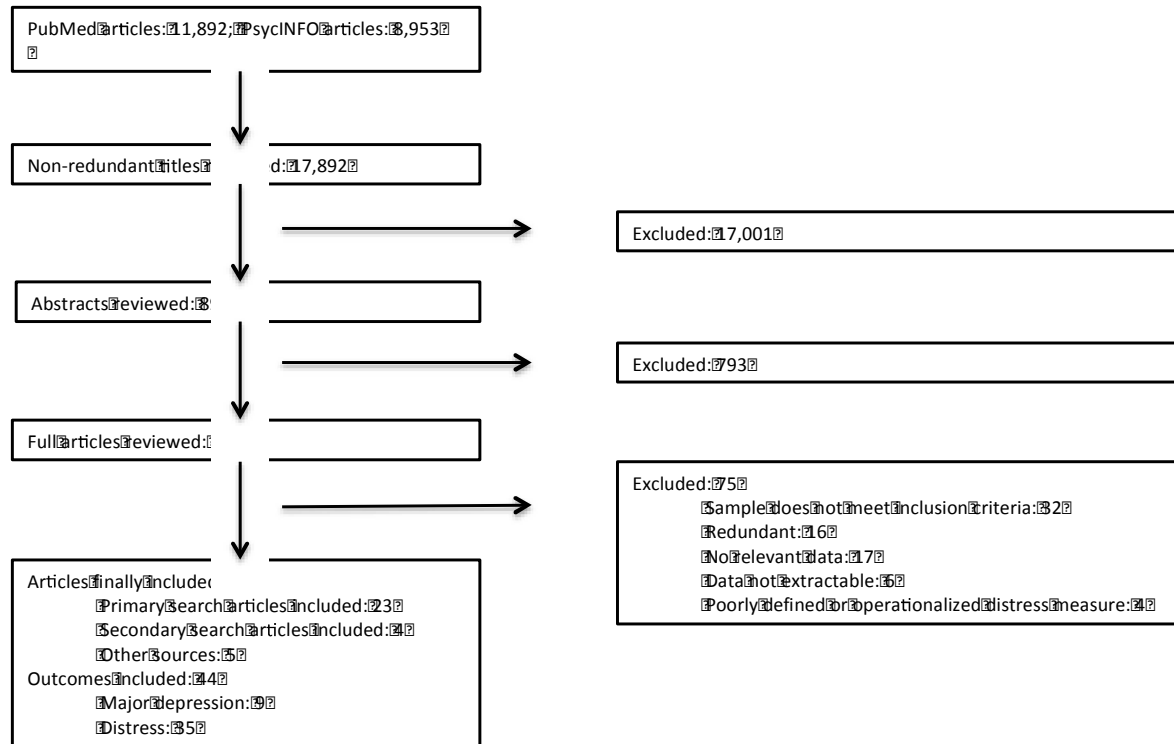[1] unweighted

Figure 2-1. Literature search flowchart

```
┌──────────────────────────────────────────────┐
│ PubMed articles: 11,892; PsycINFO articles:   │
│ 8,953                                          │
└──────────────────────────────────────────────┘
                      │
                      ▼
┌──────────────────────────────────────────────┐
│ Non-redundant titles reviewed: 17,892         │
└──────────────────────────────────────────────┘
                      │
                      │──────────────────────►  ┌─────────────────────────────────────────┐
                      │                          │ Excluded: 17,001                         │
                      ▼                          └─────────────────────────────────────────┘
┌──────────────────────────────────────────────┐
│ Abstracts reviewed: 8...                       │
└──────────────────────────────────────────────┘
                      │
                      │──────────────────────►  ┌─────────────────────────────────────────┐
                      │                          │ Excluded: 793                            │
                      ▼                          └─────────────────────────────────────────┘
┌──────────────────────────────────────────────┐
│ Full articles reviewed:                        │
└──────────────────────────────────────────────┘
                      │
                      │──────────────────────►  ┌──────────────────────────────────────────────────────────────┐
                      │                          │ Excluded: 75                                                  │
                      │                          │     Sample does not meet inclusion criteria: 32               │
                      │                          │     Redundant: 16                                             │
                      │                          │     No relevant data: 17                                      │
                      │                          │     Data not extractable: 6                                   │
                      │                          │     Poorly defined or operationalized distress measure: 4     │
                      ▼                          └──────────────────────────────────────────────────────────────┘
┌──────────────────────────────────────────────┐
│ Articles finally included:                     │
│     Primary search articles included: 23       │
│     Secondary search articles included: 4      │
│     Other sources: 5                           │
│ Outcomes included: 44                          │
│     Major depression: 9                        │
│     Distress: 35                               │
└──────────────────────────────────────────────┘
```

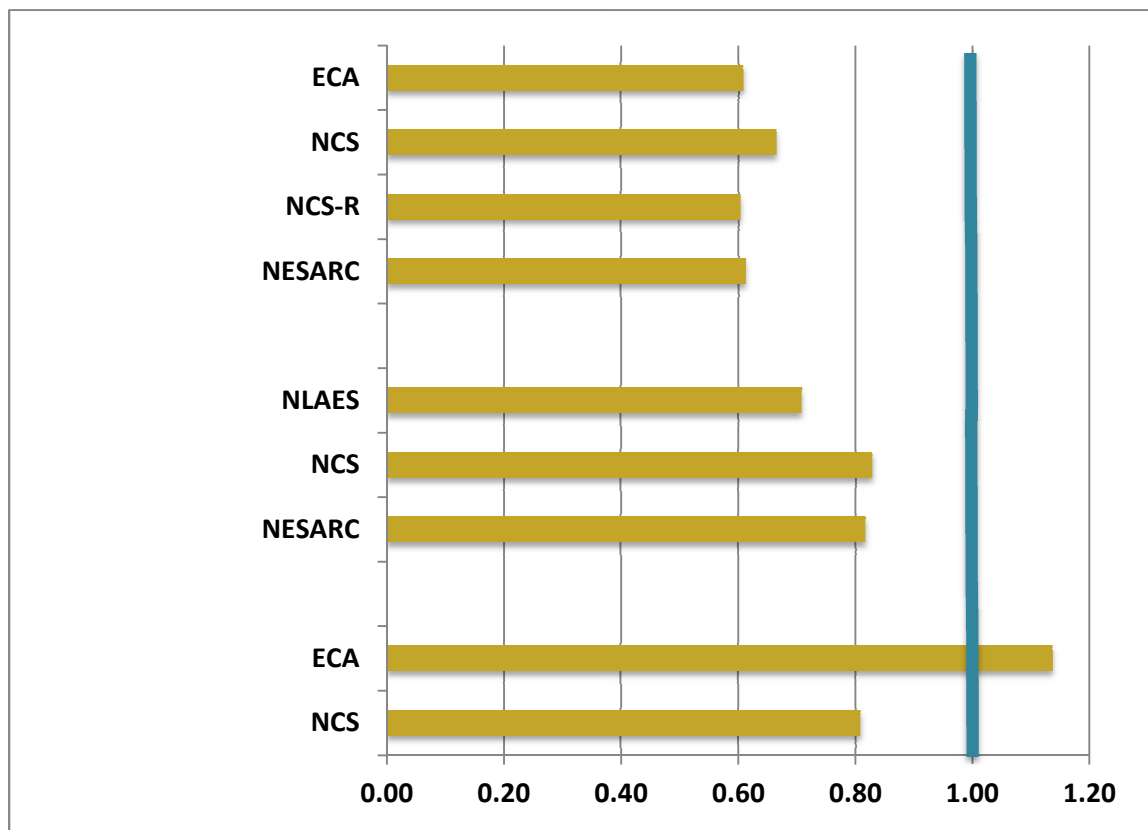**Figure 2-2. Black-White prevalence ratios of major depressive disorder**

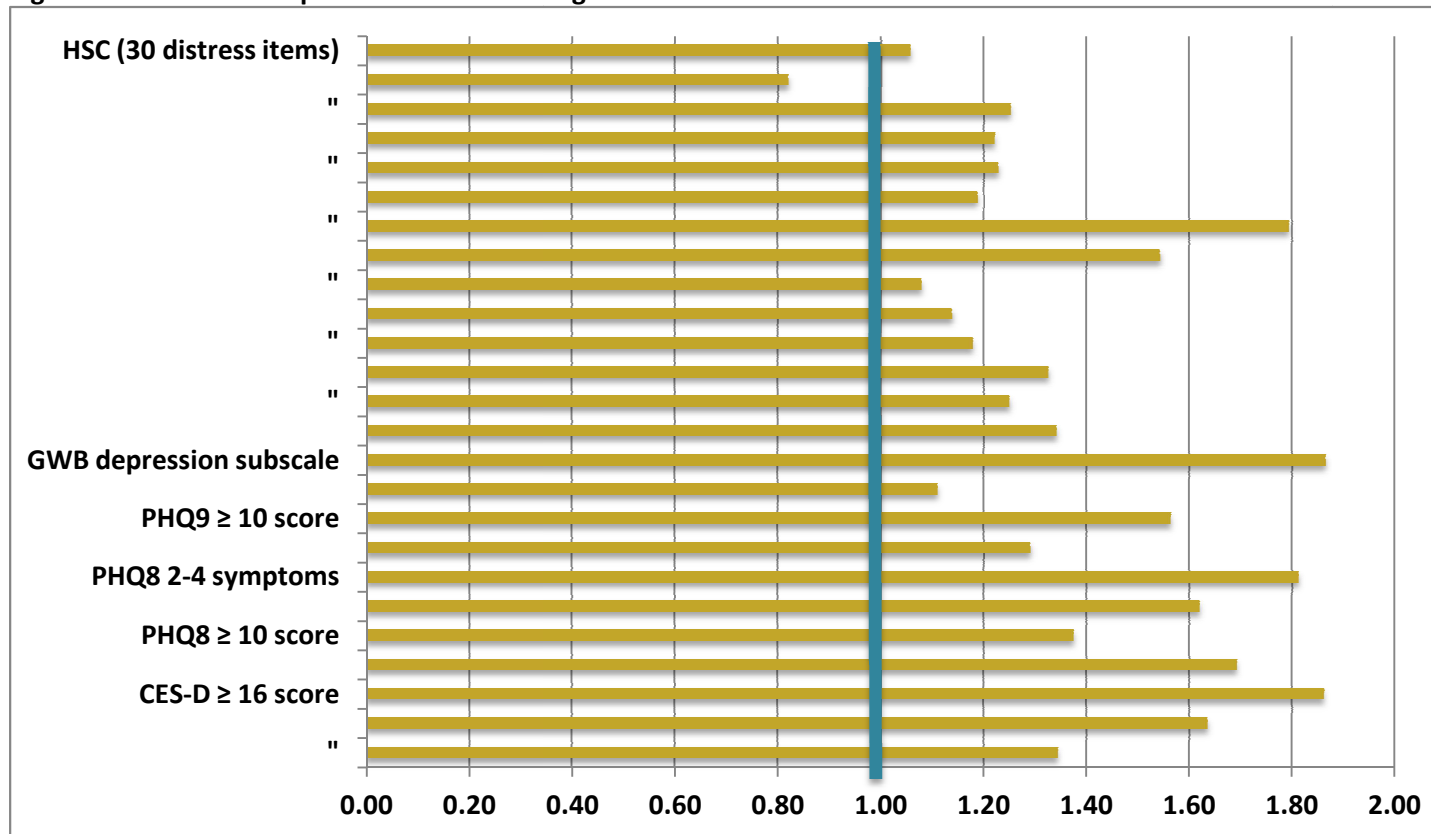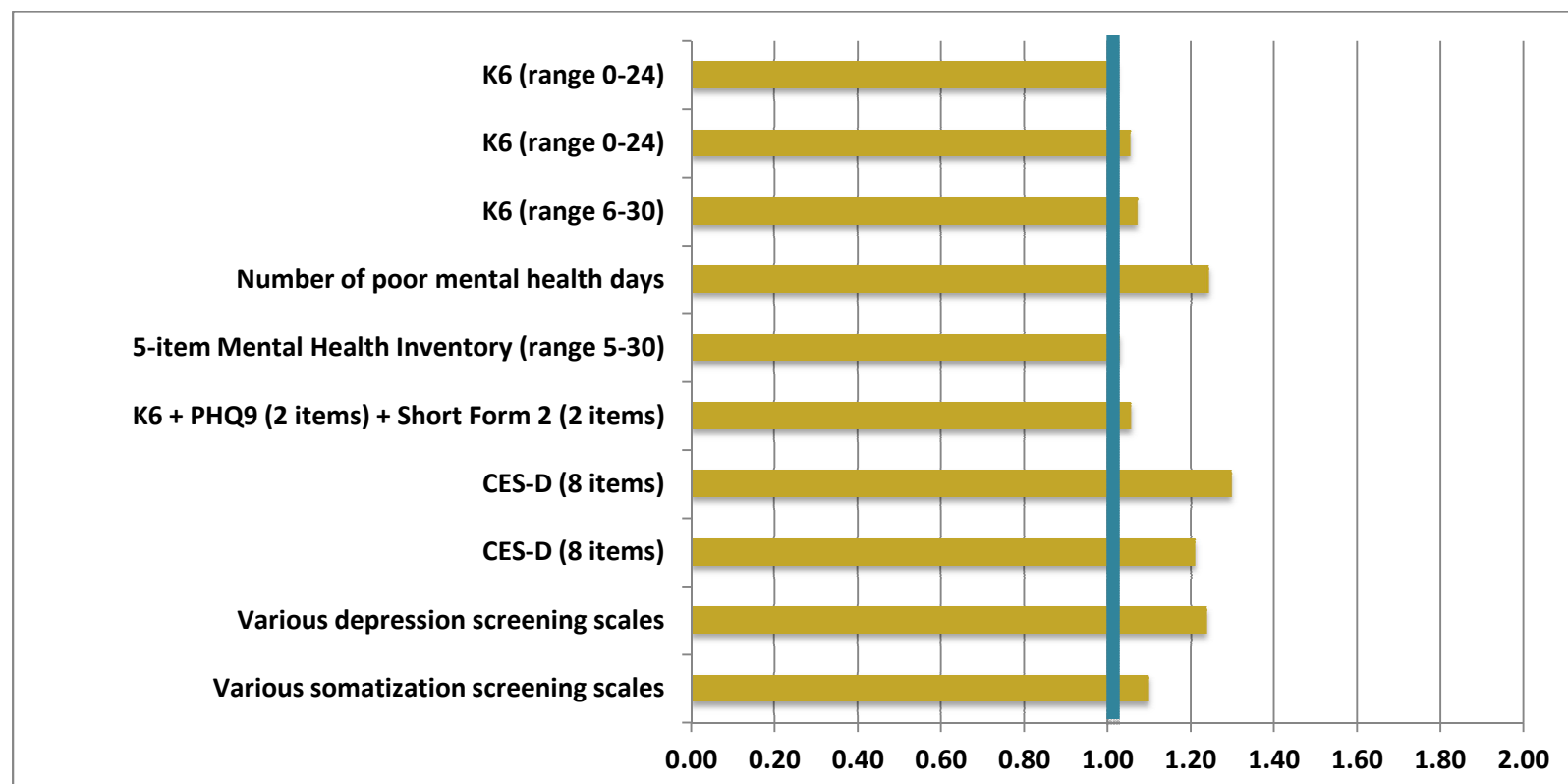**Figure 2-3. Black-White prevalence ratios of high distress**

Figure 2-4. Black-White ratios of distress means

Chapter 3: Testing a methodological explanation for the Black-White depression paradox

Introduction

In the large US psychiatric epidemiology studies using representative household samples, non-Hispanic Blacks consistently have a lower prevalence than non-Hispanic Whites (hereafter Blacks and Whites) of nearly all psychiatric disorders [32, 36, 83, 84], a finding often pronounced for major depression [35, 36, 38, 85]. Studies using similar samples also show, however, that Blacks frequently have equal or higher levels of psychological distress (hereafter distress) than Whites [4, 59, 61, 86]. This pattern of findings suggests a double paradox.

The first paradox is that the Black-White disorder findings contradict the predictions of the social stress paradigm, the dominant framework for understanding the relationship between social position and mental health [7–11]. Whether acknowledged explicitly [4] or more tacitly [2, 16, 27], this paradigm widely governs our expectations of how social status is related to mental health. The paradigm posits that disadvantaged social groups will have worse mental health outcomes than more advantaged groups because of greater stressor exposure and diminished access to coping resources [4, 87–91]. Black-White comparisons are a strong test [15] of this prediction in the American context given Blacks' disadvantaged social, political, and economic status vis-à-vis Whites, both historically and in the present [12–14, 92]. From the framework of the social stress paradigm it is therefore paradoxical that the US psychiatric epidemiology studies using nationally representative household samples produce consistent results of a lower prevalence of *any* psychiatric disorder in Blacks than Whites [36, 84], and of

most individual disorders as well [32, 36, 83, 84]. This finding is pronounced for major depressive disorder [35, 36, 38, 85].

The second paradox is the discordant Black-White disorder and distress findings in Black-White comparisons. Psychiatric disorder and psychological distress are overlapping, though distinct constructs. Both constructs define aversive mental states, are frequently measured with similar symptoms, and are phenomenologically related [18–21]. However, a distinction is often made that disorder represents dysfunction in the individual whereas distress does not assume such internal dysfunction but more often indicates the presence of stressors in the individual's environment to which the expectable response is distress [22, 23]. The constructs can be causally related when, for example, internal mental dysfunctions give rise to psychological distress or when chronic distress arising from chronic stressor exposure makes a person more vulnerable to developing internal mental dysfunctions [23]. Because of these links, we expect that those with a psychiatric disorder would score higher on distress measures than those without a disorder, and that those scoring higher on distress measures would be more likely to have a psychological disorder than those scoring lower. In fact, empirical evidence documents these associations between depression and distress [24–26]. These findings suggest that in between-group comparisons, the group with a higher prevalence of disorder should also have a higher level of distress. The apparent lack of such concordance in Black-White comparisons in the US thus constitutes the second paradox.

Resolving these paradoxes matters for two reasons. First, the Black-White disorder findings undermine our common and often tacit reliance on the social stress paradigm to predict and understand how social position relates to mental health. When a strong test fails to support the paradigm, its credibility is potentially diminished. Second, the disorder-distress discordance in Black-White

comparisons likewise potentially undermines how we conceptualize and measure perhaps the two most fundamental constructs in mental health outcomes research, disorder and distress [8, 93, 94].

Regarding the first paradox, explanations have not found robust empirical support thus far. For the most part, these explanations have focused on the major depression finding, where the lower prevalence of disorder in Blacks is pronounced. One type of explanation is substantive and starts with the assumption that the findings are valid. Examples of substantive explanations include positing greater levels in Blacks than Whites of religiosity and racial socialization [36], self-esteem [67], and social networks [66], factors thought to have protective dimensions for mental health [95–97, 97–102]. To my knowledge, only the social networks hypothesis has been tested, and the findings do not support it [66, 68]. Another example is a recent hypothesis [4, 103] proposing an interaction between race, stress, and poor health behaviors (e.g., overeating) such that at higher stressor levels, these behaviors are more protective against depression in Blacks than in Whites, while simultaneously leading to worse somatic health outcomes in Blacks. Tests of this hypothesis have had mixed results [4, 69, 70] .

A different type of explanation is artifactual and, accordingly, attributes the findings of a lower prevalence of psychiatric disorder in Blacks to methodological error. Artifactual explanations for the Black-White depression paradox include a greater tendency to misdiagnose depression as a psychotic spectrum disorder in Blacks than Whites [104–106], and differential item functioning between Blacks and Whites of the items used to diagnose major depression [27]. Differential item functioning occurs when groups have different probabilities of endorsing an item or symptom, controlling for underlying levels of the construct being measured [27]. To date, results from tests of these hypotheses explain, at best, only a small portion of the Black-White depression paradox. The prevalence of psychotic spectrum disorders [84] is too low to account for the large prevalence difference in depression, and mood

disorders more broadly, found between Blacks and Whites. Tests of differential item functioning have detected only minor differences in how the items and symptoms of the diagnostic interview for depression function between Blacks and Whites [27, 76].   A fuller discussion of differential item functioning tests is taken up below.

To date, explaining the second paradox of discordant Black-White findings between psychiatric disorder and distress has received no research attention, to my knowledge. Nevertheless, the second paradox suggests a resolution of the first paradox. That the distress findings, but not the disorder findings, cohere with social stress paradigm predictions, suggests the possibility of a methodological flaw in the diagnostic interviews used in the psychiatric epidemiology studies that does not arise in measures of distress. One clear difference between the two types of measures is that although they share symptom content, diagnostic measures often employ complex algorithms utilizing screening symptoms and exclusion criteria – features generally absent in distress measures -- that create additional opportunities for measurement error. Because the Black-White prevalence difference is pronounced for major depression, and the distress measures borrow heavily from depression symptoms [26, 107], an examination of the major depression diagnostic algorithm is a logical starting point for identifying possible artifactual problems among the diagnostic interviews.

Since the introduction of standardized criteria for diagnosing psychiatric disorders in 1980, in the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM), major depression has been formally defined by nine symptoms, and diagnosis requires endorsing at least five. At least one of the five must be either sad mood or a loss of interest or pleasure in most activities (anhedonia), hereafter referred to as screening symptoms. The remaining seven symptoms are poor concentration, feelings of worthlessness or excessive guilt, suicidality, appetite/weight change, a change in sleep

patterns, low energy, and retarded or agitated movement. The first five of these nine symptoms have been classified as psychological symptoms and the last four as somatic (i.e., physical) symptoms [19]. Thus, diagnosis entails endorsement of at least one psychological symptom but does not require endorsement of a somatic symptom. The introduction of these standardized diagnostic criteria in 1980, including the required endorsement of at least one screener, coincides with the advent of the large psychiatric epidemiology studies using representative household samples of the US. The evidence, therefore, of a Black-White depression paradox derives entirely from studies employing these diagnostic criteria.

Although depression is commonly conceptualized as comprising both psychological and somatic factors in approximately equal measure [108–112] – clearly reflected in the DSM's diagnostic symptoms -- the diagnostic algorithm advantages the psychological factor in depression by virtue of the required endorsement of a psychological screening symptom. However, the notion has been present for decades that some cultural groups experience and/or express depression more somatically than others [110, 113–118]. Ryder and colleagues [115] propose a general framework for interpreting somatic expressions of depression. First, some individuals may primarily experience, or be aware of, the somatic symptoms of their depression. Second, some individuals may have awareness of both the psychological and somatic symptoms of their depression, but the somatic are more salient to them. Third, somatization may reflect not so much the experience of depression, or the primary experience of it, but a response style in clinical or diagnostic interviews. Regarding the first two, somatization may result from culturally influenced mind-body norms, with some cultures making a weaker distinction between these realms than is normative in European cultures [114, 119]. From a more Freudian perspective, the exclusive experience of somatic symptoms, or their greater salience, may derive from ego defenses against unwanted and threatening psychological symptoms [116, 120]. With respect to response styles,

31

greater reporting of somatic symptoms may be linked with greater stigma attaching to psychological than somatic symptoms [114, 117, 118, 121]. Or, in poor resource settings, cultural norms or the conscious choice to selectively report somatic symptoms over psychological symptoms may reflect a strategy in which somatic symptoms are more likely than psychological symptoms to secure treatment [120]. Importantly, each of these manifestations of somatization would entail, though for different reasons, greater reporting of somatic symptoms in diagnostic interviews.

Also a recurring theme in the literature for several decades, often expressed by clinicians, is that Blacks in the US experience or express depression more somatically than Whites [78–80, 122]. Explanations for this alleged difference have included the following: "The denial of natural impulses and feeling, forced on blacks by racism, has created in them those symptoms that may not be representative of the typical white depressed patient. Instead, neurotic depressions are frequently manifested through somatic complaints" (p. 99) [79]. In contrast, Sleath et al [123] found evidence that clinicians may interpret the expression of emotions differently between Black and White patients, leading then to different psychotropic prescribing practices and suggesting the possibility of biases in clinicians' race-based observations. However, one need only consider how cultural differences, stemming from contrasting historical experiences, could shape the differential experience and expression of psychopathology [124, 125] to posit the possibility of Black-White differences in how depression is expressed.

Indeed, potential Black-White differences in how depression is expressed have been tested [27, 76, 111], with little evidence of greater depression somatization in Blacks than Whites [27, 76]. However, these findings have limited implications for the Black-White depression paradox. They rely on samples that have screened into the full depression interview by endorsing a screening symptom,

thereby excluding those whose more somatic expression of depression inhibits endorsing a screener. To the extent this latter group is disproportionately Black, all else being equal, depression would be disproportionately undercounted in Blacks.

If it were found that groups differed in their expression of depression on a somatic-psychological continuum, with some groups expressing depression more somatically and others expressing it more psychologically, then the DSM-defined major depression diagnostic algorithm would bias against the more somatic group, by virtue of the screening symptoms. An alternative diagnostic algorithm that is less biased against somatic manifestations of depression would forego the required screening symptoms and simply require endorsement of *any* five or more of the nine symptoms, while retaining all other elements of the DSM diagnostic algorithm. If Blacks do in fact express depression more somatically than Whites, the screening symptom requirement could lead to a more pronounced underestimate of major depression in Blacks than in Whites. The goal of the present study is to examine these questions.

Hypotheses

The foregoing discussion leads to the following hypotheses:

1. Blacks express major depression more somatically than Whites, controlling for somatic health;

2. Blacks are less likely than Whites, controlling for underlying levels of depression, to endorse the screening symptoms and greater somatization explains this lower likelihood;

3. Creating an alternative diagnostic algorithm for major depression by eliminating the screening symptom requirement, but retaining all other criteria, renders major depression

33

more prevalent in Blacks than Whites (strong version), or narrows the prevalence gap (weak version).

Control for somatic health when testing the first hypothesis is necessary to rule out the alternative explanation that Blacks endorse more somatic symptoms than Whites because of worse somatic health. Although the experience of poor somatic health is a cause of major depression, it could also, independent of depression, cause some of the somatic symptoms of depression (e.g., poor sleep or fatigue). Control for underlying depression when testing the second hypothesis is necessary to rule out the alternative explanation that Blacks are less likely than Whites to endorse the screening symptoms simply due to lower levels of depression rather than because of greater somatization.

A good test of these hypotheses – and better than previous tests of related hypotheses [27, 76, 111] -- would use data from a diagnostic interview of major depression in which information was available on all nine diagnostic symptoms from all participants in a nationally representative US sample. This would permit measuring the somatic expression of depression and the alternative diagnostic algorithm in a sample whose formation is not contingent on endorsing a psychological screening symptom. To my knowledge, the only psychiatric epidemiology study using a representative sample of the US population that collected responses on all nine major depression symptoms in the full sample is the National Longitudinal Alcohol Epidemiologic Survey (NLAES).

Methods

Study Sample and Design

This study's hypotheses are tested in data from NLAES. The National Institute of Alcohol Abuse

and Alcoholism sponsored and designed the study, which was conducted in 1991-1992 using a nationally

representative household sample of 42,862 English-speaking adults 18 and older in the 48 contiguous

states and the District of Columbia. Blacks and persons aged 18-29 were oversampled. Professional

non-clinician interviewers from the U.S. Census Bureau conducted face-to-face interviews with

participants in their homes, using the Alcohol Use Disorders and Associated Disabilities Interview

Schedule, based on the diagnostic criteria of the fourth edition of the DSM (AUDASIS-IV). The household

response rate was 91.9 percent and the person response rate was 97.4 percent [126].

Measures

Race. Those who self-reported being non-Hispanic and Black (unweighted n = 5,955; weighted percent =

12.84%) or non-Hispanic and White (unweighted n = 31,938; weighted percent = 87.16%) comprise the

sample for this study.

Age. Four age groups were created based on self-report and distributed as follows in the weighted

sample: 18-24: Blacks = 17.3%, Whites = 12.4%; 25-44: Blacks = 48.4%, Whites = 42.9%; 45-64 Blacks =

17.1%, Whites = 26.6%; ≥ 65: Blacks = 11.7%, Whites = 18.1% (χ^2 = 320.14, *df* = 3, *p* < 0.0001).   Age is

adjusted for in all analyses because of its appreciably different distribution in Blacks and Whites,

statistically significant association with screener endorsement (χ^2 = 719.67, *df* = 3, *p* < 0.0001), and

because it is not a mediator of interest in the relationship between race and mental health.

Sex. Sex was measured through self-report and is adjusted for in all analyses because of the appreciably

different sex distributions in the Black and White weighted samples (Black female = 55.3%; White female

35

= 51.9%; χ^2 = 30.77, df = 1, $p < 0.0001$), its statistically significant association with screener endorsement

(χ^2 = 279.68, df = 1, $p < 0.0001$), and because, like age, it is not a mediator of interest in the race-mental

health relationship.

Major depressive episode. The interview for major depressive episode provides all of the data for the

mental health measures in this study (underlying depression, somatization, screening symptom

endorsement, and major depressive episode). Major depressive episode itself is the outcome of the

third hypothesis and its measurement is described in detail here and the derivative measures (e.g.,

somatization) are described below. A major depressive episode is a prerequisite for the mutually

exclusive diagnoses of major depressive disorder and bipolar disorder. Individuals satisfying diagnostic

criteria for both a major depressive episode and a lifetime manic episode are diagnosed with bipolar

disorder; those meeting criteria for only a major depressive episode receive a major depressive disorder

diagnosis. Since a large majority of those with a major depressive episode receive a unipolar depression

diagnosis rather than the bipolar diagnosis [23], and because NLAES did not assess manic episodes, the

prevalence of major depressive episode serves in this study as a proxy for major depressive disorder.

The term "major depression" is used throughout to indicate "major depressive disorder," for which

major depressive episode serves as a proxy in this study.

The DSM-IV criteria for a major depressive episode – unevenly applied in NLAES, as will be

shown below -- are as follows. Individuals must endorse at least five of the nine depression symptoms,

at least one of which must be sad mood or anhedonia. All symptoms must co-occur over at least a two-

week period and occur during most days, for most of the day. However, putative cases are excluded if

any of the following sets of exclusion criteria are met: 1) the symptoms are better explained as the

direct physiological effect of a medical condition or the use of or withdrawal from medication, alcohol or

drugs; 2) the symptoms are better accounted for by bereavement within two months of the loss of a loved one, unless the bereavement is also marked by any of the following: pronounced functional impairment, psychotic symptoms, psychomotor retardation, suicidal ideation, or morbid preoccupation with worthlessness; 3) a lack of clinically significant distress or impairment in social, occupational, or other important domains of functioning; and 4) the symptoms occur as part of a mixed episode (i.e., an episode with manic features).

Major depressive episode is the only non-substance use psychiatric condition ascertained in NLAES. It was assessed using the AUDASIS-IV. This is a fully structured interview conducted by trained lay interviewers. Test-retest reliability of the AUDADIS-IV major depression interview has ranged from κ=0.65 to κ=0.73 [127–129].   Clinical reappraisal studies of major depressive episode and major depressive disorder using the AUDADIS-IV have had agreement in the κ=0.64-0.68 range [130]. Evidence from convergent validity studies was also good and is described elsewhere [38, 129, 131, 132].

Lifetime and 12-month major depressive episodes were measured in separate modules in the interview (Figure).   The lifetime module first determines the occurrence of the two screening symptoms by asking participants if in their lifetime they have ever experienced a two-week or longer period when they 1) felt sad, blue, depressed, or down most of the time, or 2) did not care about or enjoy "the usual things."   The physical illness and bereavement exclusions are enacted by asking if all of these periods occurred "when physically ill, getting over being ill, or just after someone close to you died." When participants identified two or more periods *not* excluded because of illness or bereavement, they were asked to identify either their worst (*n* = 2,573; weighted = 87.3%) or most recent (*n* = 384; weighted = 12.7%) episode. They were next asked about the other seven diagnostic symptoms co-occurring for two weeks or longer with the focal episode of sadness or anhedonia that did not occur during a period of

physical illness or bereavement. One to four questions were asked to ascertain each of the seven symptoms and endorsing any one counted as endorsing the symptom. The clinical significance exclusion criteria was operationalized in two questions asking about concurrent problems getting along with others and getting done what needed to be done. If neither item was endorsed, the case was excluded. The medication and substance use and withdrawal exclusion criteria were implemented over a series of questions at the end of the lifetime module. Manic and mixed episode exclusion criteria were not operationalized in the NLAES lifetime depression module.

In the past-year module, participants were asked about the occurrence over the past 12 months of all nine diagnostic symptoms for depressive episode, regardless of whether or not the screening symptoms were endorsed. As shown in the Figure, all participants are included in this module except those whose focal period identified in the lifetime module occurred in the past 12 months (n = 2,451, weighted = 6%), which rendered the past-year module redundant. The same symptom questions are implemented in the past-year module as in the lifetime module except that each was prefaced with "At any time in the last 12 months." Concurrence of symptoms was determined by asking of all those endorsing a screening symptom and at least three other symptoms how many periods lasting two weeks or longer in the past year they had a screening symptom "and also had some of the other experiences." The physical illness and bereavement exclusions are implemented by next asking all those endorsing symptom concurrence in the previous question whether or not all periods in the last 12 months occurred "when ill, getting over illness, or after death of someone close." Unlike the lifetime module, neither the clinical significance nor the medication, alcohol and substance use exclusion criteria were implemented in the past-year module. Like the lifetime module, the manic and mixed episode exclusion criteria were not applied in the last-year module.

Together the two modules yield data on all nine major depression symptoms over the past 12 months from all participants. From this, the outcome of the third hypothesis – past year major depressive episode measured using both the conventional and alternative algorithms -- can be tested in an unbiased sample with respect to the somatic expression of depression.

Major depression screening symptom endorsement. This is a dichotomous variable defined by whether or not participants endorsed either sad mood or anhedonia occurring over a two-week or longer period in the last 12 months.

Underlying depression. This factor was captured as a summary score of the seven non-screening symptoms endorsed in the major depressive episode interview assessing the last 12 months and has a range from 0 to 7.

Somatization. This construct was operationalized as a difference score between the number of somatic symptoms endorsed (range 0 – 4) and the number of psychological symptoms endorsed (range 0 – 3) in the major depressive episode interview assessing the last 12 months. The screening symptoms were excluded because they are the dependent variable in the second hypothesis. Scores can range from -3 to +4. Thus, a person endorsing all psychological symptoms and no somatic symptoms would have a score of -3 and a person endorsing all four somatic symptoms and no psychological symptoms would have a score of +4.

Two-thirds of the weighted sample endorsed no depression symptoms at all, including neither screener symptom, and therefore had a somatization score of "0." Thus, those providing essentially no information on depressive symptomatology "flood" the relationship – tested in the second hypothesis --

between somatization and screener endorsement. Lumping those providing no information with those endorsing an equal number of somatic and psychological symptoms (and who therefore also have a somatization score of "0") leads to an artifactual non-linear v-shaped relationship between these variables since a large majority of those with a "0" somatization score (i.e., in the approximate middle of the score range) also did not endorse a screener. To minimize these effects, the primary analyses in this study are conducted only among those endorsing at least one of the nine depression symptoms (n = 12,574). In this smaller sample, everyone provides information on somatization, thereby reducing or eliminating artifact from the somatization-screener endorsement relationship. Secondary sensitivity analyses were conducted using the full sample to examine whether the smaller sample used in the primary analyses produces noticeably discrepant findings. To account for the non-linear somatization-screener endorsement relationship caused by those providing no data on depressive symptomatology in this secondary analysis, somatization is modeled with five categories as shown in Table S1. The five categories were chosen to isolate those with "0" somatization scores (i.e., the middle of the five categories) and those at the low and high ends of somatization (i.e., the lowest and highest of the five categories).

Somatic illness. This measure was a sum of up to 23 conditions that participants identified as "causing problems in the last 12 months." (Stomach ulcer, enlarged liver, jaundice, cirrhosis of the liver, hepatitis, "some other liver disease," "high blood cholesterol, high blood fat, or high lipid content," gastritis, convulsions or epilepsy, hardening of the arteries or arteriosclerosis, high blood pressure, chest pain or angina, rapid heartbeat or tachycardia, heart attack or myocardial infarction, "other heart disease," stroke or cerebrovascular disease, emphysema, "arthritis, osteoporosis or other joint or bone disease," vitamin deficiencies or anemia, pancreatitis or any disease of the pancreas, cancer, and "any other physical health problem.")

Analytic Strategy

All analyses were conducted using SAS software's (version 9.3, SAS Institute, Inc., Cary, NC) survey procedures to account for NLAES' survey weights and complex sampling design. Accordingly, all results reported below are weighted to the US population and standard errors account for the multi-strata sampling plan. As described in the discussion of the somatization measure above, the primary analyses were conducted among those endorsing at least one of the nine depression symptoms in the past 12 months and secondary analyses were conducted in the full sample. All analyses adjust for age and sex, either through standardization when means or probabilities are compared, or in multivariable logistic regression analyses. Mean differences between Blacks and Whites in somatic health, underlying depression, somatization, and underlying dimensions of somatization were tested using SURVEYREG procedures which allowed for multivariable adjustments and age- and sex-standardizing (to the White distribution of these variables) using the *estimate* procedure. Significance tests using t-tests of these mean and probability comparisons were conducted on *openepi.com*.

The second hypothesis was tested using the SURVEYLOGISTIC procedure and adjusts for underlying depression, age, sex, and somatization.

The third hypothesis entails a Black-White comparison of major depressive episode prevalence in the full sample using conventional and alternative diagnostic algorithms. The full sample is used in testing this hypothesis because no adjustments are made for somatization. The SURVEYMEANS procedure was used to generate prevalence figures and the SURVEYLOGISTIC procedure to produce

41

odds ratios. These procedures were employed in the weighted unadjusted sample and in the weighted age- and sex-adjusted sample.

In the secondary analysis of the first hypothesis, where somatization is modeled as a categorical rather than continuous outcome, polytomous regression was employed in the SURVEYLOGISTIC procedure using a glogit link function.

Results

This study examined a methodological explanation for the consistent finding of a lower prevalence of major depression in Blacks than Whites. In the full NLAES sample, adjusting for age and sex, Blacks had a lower prevalence of past-year major depressive episode than Whites (Black prevalence = 2.03%, White prevalence = 3.63%; t = 4.14, p < 0.0001; OR = 0.60, 95% confidence interval (CI), 0.53 − 0.67). Thus, as in the three other psychiatric epidemiology studies using a representative sample of the US household population [35, 36, 38], the Black-White depression paradox was present in the NLAES sample as well.   Results from testing the three hypotheses are presented below, first from the primary sample and then from the secondary sample. All analyses were based on a sample weighted to reflect the US population, were age- and sex-adjusted, and drew only from past-year symptom reporting.

Primary analysis

The first hypothesis was that Blacks express depression more somatically than Whites, adjusting for somatic health. The results from testing this hypothesis are presented in Table 1 where somatization

means are compared. Blacks had a statistically significantly higher mean somatization score than Whites, which supports the first hypothesis. Because Blacks and Whites did not differ in their somatic health means, which are shown in the second row of results in Table 1, there was no need to adjust for this variable when testing the first hypothesis.

Table 1 also shows results of Black-White comparisons on the underlying symptom structure of the somatization measure. Among the four somatic symptoms, Blacks were statistically significantly more likely than Whites to endorse weight/appetite change, but Whites were statistically significantly more likely than Blacks to endorse low energy. Among the five psychological symptoms, Whites were more likely than Blacks to endorse each, although none of these differences was statistically significant. In a composite measure summing probabilities across the four somatic symptoms, there was no meaningful difference between Blacks and Whites. However, in the composite measure of the five psychological symptoms, Whites had a statistically significant higher score than Blacks, both when excluding and including the screening symptoms. These results suggest that Blacks' higher somatization scores were driven by the weight/appetite symptom and Whites' slightly greater tendency to endorse psychological symptoms, and not by a greater inclination in Blacks than Whites to endorse somatic symptoms generally.

The second hypothesis was that Blacks will have a lower probability than Whites of endorsing a screening symptom, adjusting for underlying depression, and that this lower probability is explained by greater somatization in Blacks. The results from logistic regression did not support the hypothesis (Table 2, models 1 and 2).   Model 1 shows that Blacks had effectively the same odds as Whites of endorsing at least one past-year screening symptom, adjusting for underlying depression, age, and sex. Model 2 shows that the Black-White odds ratio for endorsing a screener remains essentially unchanged

after adding somatization to the model. Logistic regression is typically used only for rare dichotomous outcomes (usually defined as a prevalence < 10%), in which case the odds ratio more closely approximates the more interpretable risk ratio. However, odds ratios only run the risk of dramatically overstating relative risks when the outcome is common *and* the effect size of the variable of interest is large [133]. In the present analysis, although screener endorsement prevalence exceeds 10 percent, the effect size of race on the outcome is small, in which case logistic regression does not meaningfully overstate the risk ratio.

The third hypothesis was that applying an alternative diagnostic algorithm that eliminates the screener requirement but retains all other diagnostic criteria would either make the Black prevalence of major depression higher than the White prevalence (strong version) or narrow the Black-White prevalence gap (weak version). The results from testing this hypothesis supported the weak version and are shown in Table 3. Past-year major depressive episode prevalences were compared in Blacks and Whites using the conventional and alternative algorithms in an unstandardized sample in the top panel and in an age- and sex-standardized sample in the bottom panel. The prevalence increased in both Blacks and Whites, and in both samples, when applying the alternative algorithm, but the prevalence increase was greater in Blacks than in Whites. This is most clearly seen in the odds ratios shown on the right side of Table 3. In the unstandardized sample, the Black-White odds ratio narrowed from 0.70 using the conventional algorithm to 0.81 using the unconventional algorithm. In the standardized sample, the odds ratio narrowed from 0.60 to 0.71. In both samples, the 95 percent confidence intervals around these odds ratios stayed well below 1.0.

Curiously, these results appear to be inconsistent with the results from testing the second hypothesis. Specifically, the second hypothesis results showed that there was no difference between

Blacks and Whites in the likelihood of endorsing screening symptoms, adjusting for age, sex, and underlying depression, and further adjusting for somatization did not alter this. Accordingly, removing the screening symptom requirement from the diagnostic algorithm should have had no effect on the Black-White depression odds ratio. That it *did* have an effect in the third hypothesis results is therefore unexpected. In other words, results from the second hypothesis ruled out screener endorsement as a mediator of the race-depression relationship, yet evidence from testing the third hypothesis supported mediation. This apparent contradiction between results is discussed in the Discussion section.

Secondary analysis

Results from the secondary analysis, conducted in the full sample, were generally consistent with those from the primary analysis. As discussed above, the full sample was not used in the primary analysis because two-thirds of the full sample endorsed no past-year depression symptoms and therefore contributed no information on somatization. However, to test the robustness of the findings from the smaller, primary sample, the same hypotheses were tested in the full sample but with somatization modeled categorically in most analyses, for reasons here explained. Table S1 shows how somatization scores and frequencies were arrayed across the five somatization categories in this sample, ranging from Low (those with somatization scores of -3 and -2) to High (those with somatization scores of 2, 3, and 4). Those with somatization scores of 0 formed by far the largest category and of these, 91 percent in the weighted sample endorsed no symptoms at all, including neither screening symptom. This would induce an artificially steep downward slope in the somatization-screener endorsement relationship if somatization were modeled continuously in this sample, since most in the sample would have a score at the midpoint of the somatization range and have an extremely low probability of

endorsing a screener.   By modeling somatization categorically, the anticipated v-shaped relationship between somatization and screener endorsement can be captured more precisely.

Results from testing the first hypothesis – that Blacks somatize depression more than Whites, adjusting for somatic health -- in the secondary sample are shown in Tables S2 and S3. As in the primary sample, this hypothesis was supported. Table S2 shows that Blacks had a statistically significantly higher mean somatization score than Whites when somatization was modeled continuously (it was modeled continuously in Table S2 for descriptive purposes; in Tables S3 and S4 it was modeled categorically). In this sample, the mean number of somatic health conditions was statistically significantly higher in Blacks than Whites (Table S2) and therefore the somatization mean scores adjusted for somatic health. Table S3 shows the results of comparing Blacks and Whites on somatization when modeled categorically, adjusting for age, sex, and somatic health. These polytomous regression results modeled the odds of being in each of the four higher somatization categories compared to the odds of being in the lowest category. Looking at the race variable, we see that Blacks had statistically significantly higher odds than Whites of being in the highest somatization category versus the lowest category. Blacks also had higher odds than Whites of being in the intermediate somatization categories versus the lowest category, but none of these differences was statistically significant.

Returning to Table S2 to examine the dimensions underlying somatization when it was measured continuously, the results paralleled those from the primary analysis except that none of the differences in the secondary analysis was statistically significant. The absence of statistically significant Black-White differences in the full sample on the underlying dimensions of the somatization measure presumably was due to the fact that two-thirds of this sample (in both Blacks and Whites) endorsed no

symptoms at all and the lack of Black-White differences in these participants statistically overwhelm whatever differences arose in the remaining sample.

Results from testing the second hypothesis – that Blacks will have a lower probability of endorsing a screening symptom than Whites and this will be explained by somatization -- in the full sample (Table S4) mirrored those from the primary sample. In model 1, Blacks and Whites had the same odds of endorsing a screening symptom, adjusting for underlying depression, age, and sex. Adding somatization (modeled categorically) and somatic health to the model (model 2) had no effect on the Black-White odds ratio.

Summarizing across the primary and secondary analyses, Blacks had statistically significantly higher somatization scores than Whites in both analyses, which supported the first hypothesis. However, there was no evidence that this difference constrained Blacks' likelihood of endorsing a screening symptom, a finding that did not support the second hypothesis. And finally, removing the screening symptom requirement from the diagnostic algorithm narrowed, but did not close or reverse, the major depression prevalence gap between Blacks and Whites, which supported the weaker version of the third hypothesis. As noted above, the results from testing the second and third hypotheses appeared to conflict, a question that will be examined below.

Discussion

This study tested the idea, frequently found in the literature, that Blacks express depression more somatically than Whites and that this constrains them from endorsing either of the psychological

screening symptoms required for a diagnosis of major depressive disorder, which in turn explains their lower prevalence of the disorder. Three hypotheses were tested in data from NLAES. All analyses adjusted for differences in the age and sex distributions between Blacks and Whites, were weighted to the US population, and were based on past-year symptom reporting.

Although results supported the hypothesis that Blacks express depression more somatically than Whites, the difference was small and is explained by one somatic symptom (weight/appetite change, which Blacks had a relatively high probability of endorsing compared with Whites) and a general tendency for Whites to endorse psychological symptoms more often than Blacks. However, results did not support the second hypothesis that Blacks would have a lower likelihood than Whites of endorsing a screening symptom, adjusting for underlying depression, and somatization would explain this. Blacks had the same odds as Whites of endorsing a screening symptom, whether or not somatization was in the model. Thus, there was no evidence that the small Black-White difference on the somatization measure suppressed Blacks' likelihood of endorsing the screening symptoms.

Results were consistent with the weaker version of the third hypothesis that removing the required endorsement of a screening symptom from the diagnostic algorithm for major depression, while retaining all other features of the algorithm, would attenuate the Blacks-White major depression odds ratio. However, if there was no Black-White difference in the probability of endorsing a screening symptom – as shown from testing the second hypothesis -- then removing this factor from the diagnostic algorithm should have had no effect on the Black-White odds ratio of depression. A possible explanation for this apparent contradiction is discussed below.

A preliminary question is why the Black-White difference in somatization had virtually no effect on the Black-White odds of endorsing a screener. One interpretation is that the somatization difference itself was small (on a seven-point scale, the absolute difference in mean somatization scores in the primary sample was 0.11) and was driven in part by the relatively large magnitude of the Black-White difference on the weight/appetite change symptom. Blacks' greater likelihood of endorsing this symptom may be only weakly related to sad mood or distress. For example, incidence of weight/appetite change may be due to somatic health conditions, weight loss regimens, or distress, but not depression.

Regarding the apparently inconsistent results from testing the second and third hypotheses, an explanation seems to lie with how the alternative diagnostic algorithm was operationalized in the past-year module of the NLAES. In the past-year module, the physical illness/bereavement exclusion question was asked only of those endorsing a screening symptom and three or more co-occurring additional depression symptoms. Therefore, those *not* endorsing a screening symptom but endorsing five or more other depression symptoms (and therefore satisfying the more lenient criterion of the alternative algorithm) were not asked the physical illness/bereavement exclusion question. Said differently, participants endorsing a screening symptom had the opportunity to be excluded from a diagnosis by also endorsing co-occurring physical illness or bereavement whereas those *not* endorsing a screening symptom did not have this opportunity. To the extent that Blacks were more likely than Whites to endorse this exclusion, then removing this exclusion from the diagnostic algorithm would have augmented the Black prevalence of depression more than the White prevalence when testing the alternative algorithm in NLAES.

A post-hoc analysis to investigate this possibility was conducted in the weighted, age- and sex-standardized sample. Among those asked the physical illness/bereavement exclusion question in the past-year module (i.e., among those who endorsed a screener and three or more co-occurring additional depression symptoms), 58 percent of Blacks compared with 40 percent of Whites endorsed the exclusion. If these proportions are applied to those who were *not* subject to this exclusion question in the past-year module but who *would have been* subject to it in a properly operationalized alternative algorithm (i.e., among those who endorsed five or more past-year symptoms, none of which was sad mood or anhedonia), the past-year prevalence of major depressive episode using the alternative algorithm reduces from 3.49 percent to 2.94 percent in Blacks and from 4.25 percent to 4.00 percent in Whites, which produces a revised odds ratio of 0.73. This revised odds ratio is close to the OR of 0.70 generated from the conventional algorithm. Thus, had the alternative algorithm been operationalized so that the only difference between it and the conventional algorithm was elimination of the screening symptom criteria, the Black-White odds ratios of the two algorithms would have been more or less the same. In short, the inconsistent second and third hypotheses results appear to largely be an artifact of how the alternative algorithm was necessarily, though imperfectly, implemented in the NLAES interview.

The broader question this post-hoc analysis poses is how to interpret the Black-White difference in the proportions endorsing the physical illness/bereavement exclusion among those asked this question. To the extent this Black-White difference obtains in diagnostic interviews for depression generally, and more specifically in the psychiatric epidemiology studies that reveal a Black-White depression paradox, the question is whether or not this difference reflects illness and bereavement exclusions that are legitimate applications of the DSM criteria. If it does, and no other sources of bias exist, then along with the other findings from this study, it suggests that the Black-White depression differential is a valid finding and arises in the diagnostic interview at the point of these exclusion criteria

rather than from symptom endorsement patterns. On the other hand, depression is misclassified if these two exclusion criteria are poorly operationalized, or if participants give invalid answers (i.e., by under- or over-reporting illness or bereavement). If depression misclassification stemming from the exclusion criteria is non-differential between Blacks and Whites, depression prevalence is either under-counted in both groups (because too many are excluded) or over-counted in both groups (because too few are excluded). Alternatively, if misclassification stemming from the exclusion criteria is *differential* by race, multiple possibilities exist depending on whether misclassification exists in one group or both, and in which direction(s) it occurs.

It is clear that the NLAES misclassified depressive episodes due to the study's broad operationalization of the physical illness/bereavement exclusion criteria. In the DSM, the medical condition and bereavement exclusions are designed to rule out, respectively, depressive episodes directly physiologically caused by a medical condition such as hypothyroidism, and those occurring within two months after the loss of a loved one. These criteria are operationalized far more broadly in the NLAES than the DSM conceptualizations warrant. The NLAES asks whether the depressive symptoms co-occurred "when physically ill, getting over being ill, or just after someone close to you died." This NLAES operationalization of the medical condition criteria makes no distinction between somatic conditions that can directly physiologically cause depressive symptoms (cases that should be excluded, according to the DSM) and somatic conditions that are etiologically independent of depression or that cause depression through a psychological mechanism such as the meaning a person makes of a somatic illness limiting their mobility (cases that should be included, according to the DSM). As the medical condition criterion is operationalized in the NLAES, these distinctions cannot be made and some participants will invalidly (per the DSM criteria specifications, rather than the NLAES operationalization of these criteria) endorse it, leading to an undercount of depression. A similar issue

51

pertains to the bereavement exclusion. The NLAES expression "just after someone died" under-specifies the DSM two-month cut-off and it is likely that participants variously interpreted this to imply a cut-off shorter or longer than two months. It is impossible to know how each of these scenarios actually impacted endorsement patterns in the NLAES, and how they may have differed by race.

Given the opportunities for misclassification due to the NLAES' operationalization of the physical illness/bereavement exclusion criteria, and given that misclassification patterns in Blacks and Whites in the NLAES based on these criteria is unknowable, no conclusions can be drawn about the legitimacy of the Black-White differential on endorsing these exclusion criteria. What is certain, however, is that in the NLAES sample, Black-White differences on symptom endorsements were nowhere near the scale of the Black-White difference on endorsing the physical illness/bereavement exclusion criteria. Moreover, it appears from the ad hoc analysis described above, that these exclusion criteria were nearly sufficient to explain the overall prevalence difference in major depressive episode between the two groups in this sample.

Left unexplored in the present analysis was how the clinical significance and medication/substance use exclusion criteria further shaped and ultimately determined the Black-White depression prevalence ratio in the NLAES sample. However, prior evidence indicates that few cases are excluded through these criteria [23, 37, 129]. Future studies should explore in this and other data sets which criteria in the depression algorithm most strongly determine the lower prevalence of major depression in Blacks than Whites. These detailed analyses could ultimately highlight more specifically where methodological problems with the depression interview may lay, or where protective and/or harmful factors might have their biggest impact in shaping the estimates of major depression prevalence in Blacks and Whites.

Thus, methodological explanations for the Black-White depression paradox remain to be developed and tested. The evidence presented here does not support the idea of a meaningful difference between Blacks and Whites on depression somatization; the small difference found did not account for any potential under-endorsement of the depression screening symptoms in Blacks.

Strengths and limitations

The primary strength of this study is that it relied on a large sample in which all participants provided information on all past-year depression symptoms. NLAES is the only large psychiatric epidemiology study using a nationally representative US sample in which all participants were asked about all nine depression symptoms, regardless of screener endorsement. This permitted an assessment of depression somatization in a sample that was unbiased by conditioning on endorsement of a psychological screening symptom.

Several limitations of this study are noted as well. First, the structure of the past-year depression module in NLAES led to a sub-optimal test of alternative algorithm. Nevertheless, a post-hoc analysis provided an estimate of the effect this may have had on Black and White depression prevalences using the alternative algorithm. Second, depression symptomatology may have been inaccurately reported, leading to misclassification at the symptom and diagnostic levels. To the extent this misclassification was differential by race, estimates of somatization, underlying depression, and major depressive episode prevalence may be under- or over-estimated in each race group. Third, the measure of somatic health relied on participants identifying conditions causing problems in the past year that in many cases require diagnosis by a medical professional. Strong evidence exists that Blacks
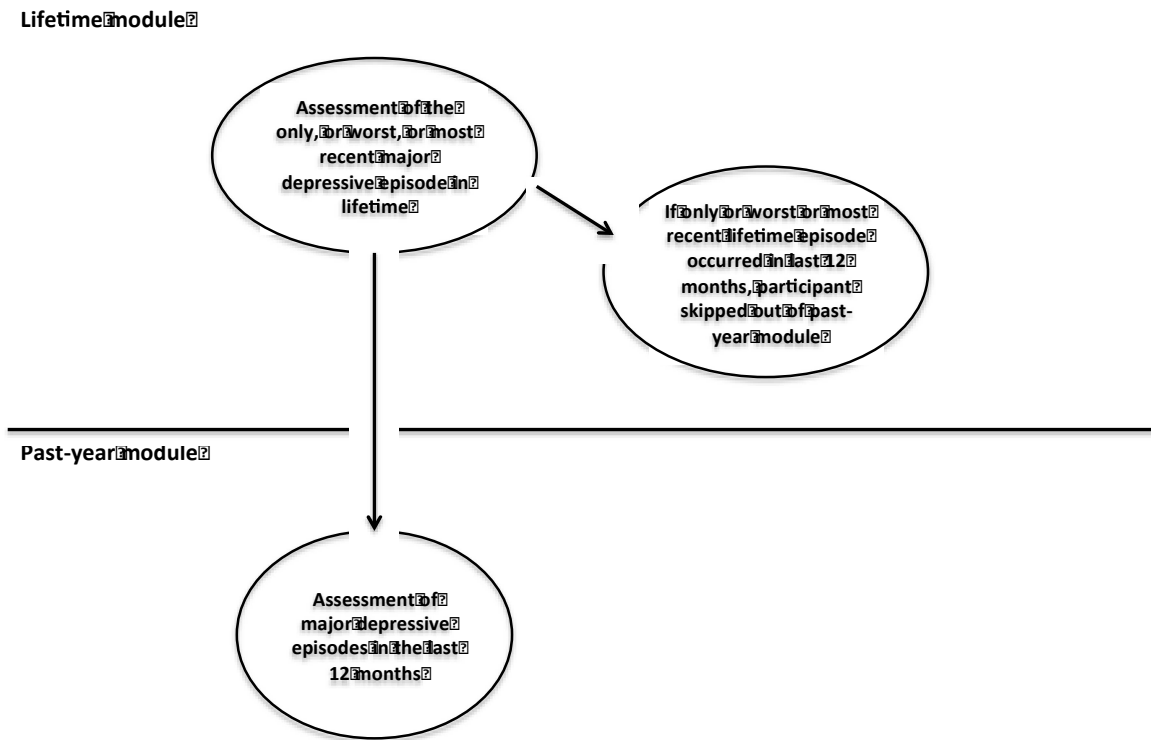
in the US have worse access to health care and receive poorer care [92, 134–138] than Whites. This may

have led to greater underestimation of poor somatic health in Blacks than Whites. In turn, this would

invalidly inflate Blacks' somatization score more than Whites' score, whether somatic health is adjusted

for or not, since some of the effect of the unreported poor somatic health is absorbed by the

somatization score. Finally, in the NLAES sample, Blacks did not have statistically significantly worse

somatic health than Whites, as this variable was measured in the study. This is inconsistent with the

large body of research documenting that Blacks have worse somatic health than Whites in the US [92,

139–141]. It could be due to disproportionate, though inadvertent, under-reporting of somatic

conditions in Blacks, as just discussed, or to the limited number of somatic conditions measured, or to

random chance in the sample. The first two reasons have the same implications for the findings: Blacks'

somatization score would be biased upward more than Whites' score. However, since there was no

evidence that the small Black-White difference on somatization, as measured, suppressed Blacks'

endorsement of screening symptoms, this possible inflation of Blacks' somatization score would have no

impact on my conclusions. On the other hand, if the somatic health of Blacks in this sample was better,

by chance, than the somatic health of adult Blacks living in households nationwide, it is not clear that

this would have impacted the conclusions. Black somatization scores may have been higher in a more

representative Black sample, in terms of somatic health, but adjusting for somatic health ostensibly

would have accounted for this.

Conclusion

In this study, no evidence was found for a meaningful difference between Blacks and Whites on

the somatization of depression. Blacks' slightly higher somatization score was driven by one somatic

symptom and a greater tendency in Whites to endorse psychological symptoms. Adjusting for Blacks'

slightly higher level than Whites of depression somatization had no impact on Blacks' likelihood of endorsing a screening symptom. On the other hand, a large difference was found between Blacks and Whites on the likelihood of endorsing the physical illness and bereavement exclusion criteria. This difference accounted for virtually all of the lower prevalence of major depression in Blacks than Whites. Future studies should conduct detailed analyses in this and other data sets of where in the diagnostic interview for depression the Black-White prevalence gap arises. Importantly, these analyses should be guided by theory regarding group differences. Such theory will make the researchers' assumptions explicit and presumably coherent, and will lead to a more efficient resolution of the Black-White depression paradox.

Figure 3-1. Major depression modules in the National Longitudinal Alcohol Epidemiologic Survey (1991-1992)

Lifetime module

Assessment of the only, or worst, or most recent major depressive episode in lifetime

If only or worst or most recent lifetime episode occurred in last 12 months, participant skipped out of past-year module

Past-year module

Assessment of major depressive episodes in the last 12 months

Table 1. Blacks' and Whites' age- and sex-standardized means and prevalences among those endorsing at least one past-year depression symptom in the National Longitudinal Alcohol Epidemiologic Survey (1991-1992)

	Black 2,122 (13.64%)[1]	White 10,452 (86.36%)	t value	p value
Somatization mean (SE), range -3 to +4	0.65 (0.04)	0.54 (0.02)	-2.70	0.01
Somatic health conditions mean (SE), range 0-22	1.06 (0.04)	1.02 (0.02)	-0.81	0.42
Underlying depression mean (SE), range 1-7	2.52 (0.05)	2.57 (0.03)	0.69	0.49
Underlying dimensions of somatization and depression				
Somatic symptom endorsement probabilities, range 0-1 (SE)				
Weight/appetite change	0.55 (0.01)	0.45 (0.01)	-4.41	<0.0001
Sleep problems	0.41 (0.01)	0.43 (0.01)	0.88	0.38
Psychomotor changes	0.25 (0.01)	0.24 (0.01)	-0.44	0.66
Low energy	0.37 (0.01)	0.42 (0.01)	2.21	0.03
Psychological symptom endoresement probabilities, range 0-1 (SE)				
Sad mood	0.34 (0.01)	0.36 (0.01)	0.88	0.38
Anhedonia	0.26 (0.01)	0.29 (0.01)	1.32	0.19
Poor concentration	0.31 (0.01)	0.32 (0.01)	0.44	0.66
Guilt/worthlessness	0.23 (0.01)	0.27 (0.01)	1.77	0.08
Suicidality	0.39 (0.02)	0.43 (0.01)	1.77	0.08
Somatic symptom probability sums, range 0-4 (SE)	1.59 (0.04)	1.55 (0.02)	-0.84	0.40
Psychological symptom probability sums, range 0-3 (SE) [2]	0.93 (0.03)	1.02 (0.01)	3.46	< 0.001
Psychological symptom probability sums, range 0-5 (SE)[3]	1.53 (0.05)	1.66 (0.02)	2.61	0.01

Note. All comparisons are based on weighted sample.

[1] unweighted *n* (weighted %)

[2] excluding screener symptoms

[3] including screener symptoms

Table 2. Logistic regression of race predicting screener endorsement, adjusting for underlying depression, age, sex, and somatization, among those endorsing at least one past-year depression symptom in the National Longitudinal Alcohol Epidemiologic Survey (1991-1992)

	Model 1			Model 2	
	OR	95% CI		OR	95% CI
Race (ref = White)	0.97	0.89 - 1.07		1.00	0.91 - 1.09
Underlying depression	2.12	2.07 - 2.18		2.15	2.10 - 2.21
Age	0.82	0.79 - 0.86		0.83	0.79 - 0.86
Sex (ref = male)	1.06	0.98 - 1.15		1.12	1.03 - 1.21
Somatization				0.78	0.76 - 0.80

Note. Both analyses based on weighted sample.

Table 3. Prevalence of major depressive episode using conventional and unconventional algorithms in the National Longitudinal Alcohol Epidemiologic Survey (1991-1992)

	Black 5,955 (12.84%)[1]	White 31,938 (87.16%)	*t* value	*p* value	OR	95% CI
Unstandardized						
conventional algorithm[2]	2.55 (0.00)	3.63 (0.00)	-3.30	0.001	0.70	0.62 - 0.78
alternative algorithm[3]	3.49 (0.00)	4.25 (0.00)	-2.04	0.04	0.81	0.74 - 0.89
Standardized[4]						
conventional algorithm	2.03 (0.00)	3.63 (0.00)	-4.14	<0.0001	0.60	0.53 - 0.67
alternative algorithm	2.98 (0.00)	4.25 (0.00)	-2.91	0.004	0.71	0.65 - 0.79

Note. All analyses used weighted sample.

[1] unweighted *n* (weighted %)

[2] applying the DSM algorithm as implemented in NLAES

[3] applying the DSM algorithm as implemented in NLAES, except the requirement of endorsing a screening symptom

[4] age- and sex-standardized

Table S1. Somatization categories in the full sample, age-
and sex-standardized, in the National Longitudinal
Alcohol Epidemiologic Survey (1991-1992)

Somatization category	Somatization scores	n (%)[1]
Low	-3, -2	326 (0.9)
Medium low	-1	2,532 (6.6)
Medium	0	27,622 (73.9)
Medium high	1	4,966 (12.7)
High	2, 3, 4	2,447 (6.0)

[1] unweighted *n* (weighted %)

Table S2. Blacks' and Whites' age-and sex-standardized means and prevalences in the full sample in the National Longitudinal Alcohol Epidemiologic Survey (1991-1992)

	Black	White		
	5,955 (12.84%)[1]	31,938 (87.16%)	*t* value	*p* value
Somatization mean (SE), range -3 to +4 [2]	0.15 (0.02)	0.11 (0.01)	-2.48	0.01
Somatic health conditions mean (SE), range 0-22	0.73 (0.03)	0.66 (0.01)	-2.08	0.04
Underlying depression mean (SE), range 1-7	0.84 (0.04)	0.83 (0.02)	-0.21	0.84
Underlying dimensions of somatization and depression				
Somatic symptom endorsement probabilities, range 0-1 (SE)				
Weight/appetite change	0.18 (0.01)	0.15 (0.01)	-1.32	0.19
Sleep problems	0.14 (0.01)	0.14 (0.01)	0.23	0.81
Psychomotor changes	0.09 (0.01)	0.08 (0.00)	-0.90	0.37
Low energy	0.13 (0.01)	0.14 (0.00)	0.95	0.34
Psychological symptom endoresement probabilities, range 0-1 (SE)				
Sad mood	0.11 (0.01)	0.12 (0.00)	0.12	0.91
Anhedonia	0.09 (0.01)	0.09 (0.00)	0.62	0.53
Poor concentration	0.11 (0.01)	0.10 (0.00)	-0.28	0.78
Guilt/worthlessness	0.08 (0.01)	0.09 (0.00)	1.60	0.11
Suicidality	0.13 (0.01)	0.14 (0.00)	0.71	0.48
Somatic symptom probability sums, range 0-4 (SE)	0.53 (0.03)	0.50 (0.02)	-0.89	0.38
Psychological symptom probability sums, range 0-3 (SE)[3]	0.31 (0.02)	0.33 (0.01)	0.71	0.48
Psychological symptom probability sums, range 0-5 (SE)[4]	0.51 (0.03)	0.54 (0.02)	0.60	0.55

Note. All comparisons are based on weighted sample.

[1] unweighted *n* (weighted %)

[2] adjusting for somatic health

[3] excluding screener symptoms

[4] including screener symptoms

61

Table S3. Polytomous regression of somatization on race, adjusting for underlying depression, age, sex, and somatic health in the full sample, in the National Longitudinal Alcohol Epidemiologic Survey (1991-1992)

	Somatization (ref = low)[1]	OR	95% CI
Race (ref = White)	medium low	1.11	0.89 - 1.38
	medium	1.14	0.93 - 1.40
	medium high	1.13	0.92 - 1.39
	high	1.40	1.11 - 1.76
Age	medium low	1.24	1.09 - 1.40
	medium	1.61	1.42 - 1.83
	medium high	1.17	1.03 - 1.32
	high	1.14	0.99 - 1.30
Sex (ref = Male)	medium low	1.30	1.07 - 1.56
	medium	1.28	1.05 - 1.55
	medium high	1.65	1.35 - 2.01
	high	2.10	1.69 - 2.60
Somatic health	medium low	0.87	0.79 - 0.95
	medium	0.70	0.65 - 0.77
	medium high	1.01	0.92 - 1.10
	high	1.12	1.02 - 1.22

Note. Analysis used weighted sample.
[1] Somatization scores in each category are: "low" (ref) (-2 and -3); "medium low" (-1); "medium" (0; which includes all those endorsing none of the seven symptoms); "medium high" (1 and 2); "high" (3 and 4).

Table S4. Logistic regression of depression screener endorsement on race, adjusting for underlying depression, age, sex, somatization, and somatic health in the full sample, in the National Longitudinal Alcohol Epidemiologic Survey (1991-1992)

		Model 1		Model 2	
		OR	95% CI	OR	95% CI
Race (ref = White)		0.99	0.89 - 1.10	1.01	0.91 -1.11
Underlying depression		3.14	3.06 - 3.22	3.11	3.02 - 3.20
Age		0.80	0.76 - 0.83	0.81	0.77 - 0.85
Sex (ref = male)		1.07	0.99 - 1.17	1.12	1.03 - 1.22
Somatization (ref = low)	medium low			0.53	0.43 - 0.65
	medium			0.23	0.18 - 0.29
	medium high			0.27	0.22 - 0.34
	high			0.21	0.17 - 0.27
Somatic health				0.99	0.95 - 1.04

Note. Both analyses based on weighted sample

Chapter 4: Testing a methodological explanation for the paradoxical Black-White findings on DEPRESSION AND distress

Introduction

An unsolved conundrum consistently appears in large, population-based studies measuring mental health in the US: Blacks are reliably shown to have a lower prevalence than Whites of most psychiatric disorders [32, 36, 83, 84], yet equal and higher levels of psychological distress [1]. These findings suggest a non-trivial double paradox.

The first paradox is that the Black-White disorder findings contradict the predictions of the social stress paradigm, the dominant framework for understanding the relationship between social position and mental health [7–11]. Whether acknowledged explicitly [4] or more tacitly [2, 16, 27], this paradigm widely governs our expectations of how social status is related to mental health. The paradigm posits that disadvantaged social groups will have worse mental health outcomes than more advantaged groups because of greater stressor exposure and diminished access to coping resources [4, 87–91]. Black-White comparisons are a strong test [15] of this prediction in the American context given Blacks' disadvantaged social, political, and economic status vis-à-vis Whites, both historically and in the present [12–14, 92]. From the framework of the social stress paradigm it is therefore paradoxical that the US psychiatric epidemiology studies using nationally representative household samples produce consistent results of a lower prevalence of *any* psychiatric disorder in Blacks than Whites [36, 84], and of most individual disorders as well [32, 36, 83, 84]. This finding is relatively pronounced for major depressive disorder [35, 36, 38, 85].

The second paradox is the discordant Black-White disorder and distress findings in Black-White comparisons. Psychiatric disorder and psychological distress are overlapping, though distinct constructs. Both constructs define aversive mental states, are frequently measured with similar symptoms, and are phenomenologically related [18–21]. However, a distinction is often made that disorder represents dysfunction in the individual whereas distress does not assume such internal dysfunction but more often indicates the presence of stressors in the individual's environment to which the expectable response is distress [22, 23]. The constructs can be causally related when, for example, internal mental dysfunctions give rise to psychological distress or when chronic distress arising from chronic stressor exposure makes a person more vulnerable to developing internal mental dysfunctions [23]. Because of these links, we expect that those with a psychiatric disorder would score higher on distress measures than those without a disorder, and that those scoring higher on distress measures would be more likely to have a psychological disorder than those scoring lower. In fact, empirical evidence documents these associations between depression and distress [24–26]. These findings suggest that in between-group comparisons, the group with a higher prevalence of disorder should also have a higher level of distress. The apparent lack of such concordance in Black-White comparisons in the US thus constitutes the second paradox.

Resolving these paradoxes matters for two reasons. First, the Black-White disorder findings undermine our common and often tacit reliance on the social stress paradigm to predict and understand how social position relates to mental health. When a strong test fails to support the paradigm, its credibility is potentially diminished. Second, the disorder-distress discordance in Black-White comparisons likewise potentially undermines how we conceptualize and measure perhaps the two most fundamental constructs in mental health outcomes research, disorder and distress [8, 93, 94].

The present study focuses on the second paradox. To date, explaining the second paradox of discordant Black-White findings between psychiatric disorder and distress has received no research attention, to my knowledge. Nevertheless, the second paradox suggests a resolution of the first paradox. That the distress findings, but not the disorder findings, cohere with social stress paradigm predictions, suggests the possibility of a methodological flaw in the diagnostic interviews used in the psychiatric epidemiology studies that does not arise in measures of distress. One clear difference between the two types of measures is that although they share symptom content, diagnostic measures often employ complex algorithms utilizing screening symptoms and exclusion criteria – features generally absent in distress measures -- that create additional opportunities for measurement error. Because the Black-White prevalence difference is pronounced for major depression, and the distress measures borrow heavily from depression symptoms [26, 107], an examination of the major depression diagnostic algorithm is a logical starting point for identifying possible artifactual problems among the diagnostic interviews.

Since the introduction of standardized criteria for diagnosing psychiatric disorders in 1980, in the third edition of the Diagnostic and Statistical Manual (DSM), major depression has been formally defined by nine symptoms, with diagnosis requiring endorsement of at least five. At least one of the five must be either sad mood or a loss of interest or pleasure in most activities (anhedonia), hereafter referred to as screening symptoms. The remaining seven symptoms are poor concentration, feelings of worthlessness or excessive guilt, suicidality, appetite/weight change, a change in sleep patterns, low energy, and retarded or agitated movement. The first five of these nine symptoms have been classified as psychological symptoms and the last four as somatic (i.e., physical) symptoms [19]. Thus, diagnosis entails endorsement of at least one psychological symptom but does not require endorsement of a

somatic symptom. The introduction of these standardized diagnostic criteria in 1980, including the required endorsement of at least one screener, coincides with the advent of the large psychiatric epidemiology studies using representative household samples of the US. The evidence, therefore, of a Black-White depression paradox derives entirely from studies employing these diagnostic criteria.

Although depression is commonly conceptualized as comprising both psychological and somatic factors in approximately equal measure [108–112] – clearly reflected in the DSM's diagnostic symptoms -- the diagnostic algorithm advantages the psychological factor in depression by virtue of the required endorsement of a psychological screening symptom. However, the notion has been present for decades that some cultural groups experience and/or express depression more somatically than others [110, 113–118].   Ryder and colleagues [115] propose a general framework for interpreting somatic expressions of depression. First, some individuals may primarily experience, or be aware of, the somatic symptoms of their depression. Second, some individuals may have awareness of both the psychological and somatic symptoms of their depression, but the somatic are more salient to them.   Third, somatization may reflect not so much the experience of depression, or the primary experience of it, but a response style in clinical or diagnostic interviews. Regarding the first two, somatization may result from culturally influenced mind-body norms, with some cultures making a weaker distinction between these realms than is normative in European cultures [114, 119]. From a more Freudian perspective, the exclusive experience of somatic symptoms, or their greater salience, may derive from ego defenses against unwanted and threatening psychological symptoms [116, 120] With respect to response styles, greater reporting of somatic symptoms may be linked with greater stigma attaching to psychological than somatic symptoms [114, 117, 118, 121]. Or, in poor resource settings, cultural norms or the conscious choice to selectively report somatic symptoms over psychological symptoms may reflect a strategy in which somatic symptoms are more likely than psychological symptoms to secure treatment

[120]. Importantly, each of these manifestations of somatization would entail, though for different reasons, greater reporting of somatic symptoms in diagnostic interviews.

Also a recurring theme in the literature for several decades, often expressed by clinicians, is that Blacks in the US experience or express depression more somatically than Whites [78–80, 122]. Explanations for this alleged difference have included the following: "The denial of natural impulses and feeling, forced on blacks by racism, has created in them those symptoms that may not be representative of the typical white depressed patient. Instead, neurotic depressions are frequently manifested through somatic complaints" (p. 99) [79]. In contrast, Sleath et al [123] found evidence that clinicians may interpret the expression of emotions differently between Black and White patients, leading then to different psychotropic prescribing practices and suggesting the possibility of biases in clinicians' race-based observations. However, one need only consider how cultural differences, stemming from contrasting historical experiences, could shape the differential experience and expression of psychopathology [124, 125] to posit the possibility of Black-White differences in how depression is expressed.

The relevance to the second paradox of a Black-White difference on the somatic dimension of depression is that the diagnostic algorithm for depression potentially biases against somatic expressions of distress in a way that measures of distress do not. Whereas a major depression diagnosis requires endorsing a psychological screening symptom, no such algorithm exists for measures of distress where all items are typically given equal weight. Thus, to the extent Blacks *do* somatize depression and other forms of distress more than Whites, measures of distress that sufficiently tap somatic symptoms will more accurately reflect their distress than major depression diagnoses will their depression.

68

Some evidence exists of greater distress and depression somatization in Blacks than Whites [27, 76, 77, 111, 142, 143], but it tends to be symptom- or item-specific, the symptoms vary across studies, and to date there is scant evidence to support a broad somatization hypothesis that Blacks endorse a spectrum of somatic symptoms more than Whites after adjusting for underlying levels of distress or depression. However, these findings have two main limitations. First, to my knowledge, no study has compared Blacks and Whites on distress and depression in the *same* sample with the goal of understanding discordant results across outcomes. Second, the studies testing differential item functioning between Blacks and Whites on depression symptoms were all conducted in samples in which participants screened into the full depression interview because they endorsed at least one screening symptom, thereby biasing the sample against those whose more somatic expressions of depression may have inhibited endorsement of a screening symptom. Differential item functioning occurs when two or more groups of interest have different probabilities of endorsing items or symptoms on a measure after controlling for underlying levels of the construct of interest. Groups may differ in their probabilities of endorsing a particular item in a measure, but this could be due to overall lower scores on the measure; differential item functioning tests forestall this explanation by adjusting for underlying levels of the measure. The present study circumvents these limitations, first, by comparing Blacks and Whites on both distress and depression in the same sample, and second, by assessing somatization using the distress measure which all participants completed (and which thereby provides a measure of somatization in an unbiased sample).

Hypotheses

The current study examines whether Blacks express distress more somatically than Whites and then tests whether Blacks are less likely to endorse depression screening symptoms, adjusting for

distress, because of greater distress somatization. The following specific hypotheses are tested:

1. Blacks will somatize the expression of distress more than Whites, adjusting for somatic health.

2. Blacks will have a lower likelihood than Whites of endorsing a screening symptom, adjusting for distress, and this will be partly (weak version) or fully (strong version) explained by Blacks' greater somatization.

Control for somatic health when testing the first hypothesis is necessary to rule out the alternative explanation that Blacks endorse more somatic symptoms than Whites because of worse somatic health. Although the experience of poor somatic health is a cause of major depression, it could also, independent of depression, cause some of the somatic symptoms of depression (e.g., poor sleep or fatigue).

Methods

Study

These hypotheses are tested in data from the National Comorbidity Survey Replication (NCS-R). This survey was conducted between 2001 and 2003 in a nationally representative sample of English-speaking civilians 18 and older living in non-institutionalized settings, and students living in campus housing who had permanent household addresses, in the 48 contiguous states. It was one of three concurrent psychiatric epidemiology studies sponsored by the Inter-University Consortium for Political

and Social Research, and collectively called the Collaborative Psychiatric Epidemiology Surveys [144]. The NCS-R response rate was 73.0 percent [145]. Professional non-clinician interviewers conducted face-to-face interviews with 9,282 participants using laptop computer-assisted personal interview (CAPI) methods. The interview was conducted in two parts during the same appointment. Part 1 consisted primarily of psychiatric diagnostic interviews and part 2 collected data on risk factors for disorder and a subset of psychiatric disorders not diagnosed in part 1 [32]. Part 2 was conducted among all those diagnosed in Part 1 with a lifetime disorder and a random selection of the remaining Part 1 participants [145]. The part 2 sample size was 5,692. The present study draws only on participants from the part 2 sample because the distress measure was administered only to them. The part 1 sample was weighted to account for differential probabilities of selection into the study and to match the US population in 2000 on socio-demographic and geographic factors. The part 2 sample was weighted to adjust for these same factors as well as differential probabilities of selection from part 1 of the study [146]. Psychiatric diagnoses were made in the NCS-R using the World Mental Health Survey Initiative Version of the World Health Organization's Composite International Diagnostic Interview (WMH-CIDI), a fully structured, lay-administered instrument based on DSM-IV criteria [145]. The WMH-CIDI was used in the NCS-R to diagnose eight different anxiety disorders, four mood disorders (including major depressive disorder), four impulse control disorders, and four substance use disorders.

Measures

Race. The study sample was comprised of part 2 participants who self-reported being non-Hispanic and Black (unweighted *n* = 717; weighted percent = 14.53%) or non-Hispanic and White (unweighted *n* = 4,180; weighted percent = 85.47%).

Age. Four age groups were created based on self-report and distributed as follows in the weighted sample: 18-24: Blacks = 20.0%, Whites = 12.7%; 25-44: Blacks = 40.2%, Whites = 34.5%; 45-64 Blacks = 28.2%, Whites = 33.4%; ≥ 65: Blacks = 11.7%, Whites = 19.5% (χ^2 = 53.39, *df* = 3, *p* < 0.0001). Age is adjusted for in all analyses because of its appreciably different distribution in Blacks and Whites, its statistically significant association with screener endorsement (χ^2 = 59.01, *df* = 3, *p* < 0.0001), and because it is not a mediator of interest in the relationship between race and mental health.

Sex. Sex was measured through self-report and is adjusted for in all analyses because of the appreciably different sex distributions in the weighted Black and White samples (Black female = 57.7%; White female = 52.5%; χ^2 = 6.71, *df* = 1, *p* = 0.01), its statistically significant association with screener endorsement (χ^2 = 17.43, *df* = 1, *p* < 0.0001), and because, like age, it is not a mediator of interest in the race-mental health relationship.

Major depression screening symptom endorsement. This is a dichotomous variable derived from the major depression interview and defined by whether or not participants endorsed either sad mood or anhedonia occurring over a two-week or longer period in the last 12 months. The major depression interview is a component of the WMH-CIDI diagnostic interview and was administered to all part 1 participants. A blind clinical re-appraisal study comparing NCS-R CIDI past-year major depressive episode diagnoses with Structured Clinical Interview for DSM-IV (SCID) major depressive episode diagnoses produced a *Kappa* of .40 (95% Confidence Interval (CI), 0.20 - 0.60) [145]. In the National Survey of American Life, conducted simultaneously with the NCS-R, the same comparison yielded *Kappas* of 0.43 (95% CI, 0.26 – 0.59) in African-Americans and 0.27 (-0.13 – 0.67) in Whites [67]. These are considered "fair" *Kappas* at best [67].

Participants skipped out of the past-year depression module if they had previously indicated in the interview that they had never in their lifetime experienced a screening symptom lasting two weeks or longer, for at least one hour each day, while concurrently experiencing at least one additional, non-screen symptom nearly every day during the two-week or longer period. Everyone not skipping out was asked at the outset of the past-year depression module: "Did you have an episode of being (sad/or/discouraged/or/uninterested) [a participant's prior responses indicated which of these terms to use] with some of the other problems (IF R[espondent] CAN READ: you mentioned on pages 4-5/IF R CANNOT READ: we just reviewed) lasting two weeks or longer at any time in the past 12 months?" Those endorsing this were counted as a "yes" (n = 700; weighted percent = 8.8%). Those responding "no," and those not asked the question because of previous skip-outs were counted as a "no" (n = 4,197; weighted percent = 91.2%). "The other problems" mentioned in the question refers to the seven non-screen depression symptoms.

Psychological distress. The NCS-R used the K10 instrument to measure non-specific psychological distress. It is a widely used, and perhaps the preeminent [19], measure of distress, originally developed as a brief screen for serious mental illness in the US National Health Interview Survey and designed to maximize sensitivity at the more severe range of psychiatric disorder [20]. As implemented in the NCS-R, the K10 asked participants to consider the month in the last year when they were at their worst "emotionally, in terms of being anxious, depressed, or emotionally stressed." Participants were then asked the extent to which they experienced 10 symptoms during the focal month. Response options for each item are "all of the time," "most of the time," "some of the time," "little of the time," and "none of the time," and were scored from 1 to 5, respectively. For the present study, responses were reverse coded and anchored at 0 so that higher scores indicated higher distress levels. Scores are summed across items, creating a composite score that in the present study could range from 0 to 40. The ten

distress symptoms are: "tired for no reason," "felt nervous," "felt so nervous nothing could calm you down," "felt hopeless," "was restless/fidgety," "was so restless you couldn't sit still," "felt depressed," "felt so depressed nothing could cheer you up," "felt everything was an effort," and "felt worthless." The third, sixth, and eighth of these questions were asked only of those endorsing the prior item.

In initial validity studies in a general population sample, using 12-month major depression diagnoses from the Structured Clinical Interview DSM-IV (SCID) to establish caseness, the area under the receiver operator characteristic curve for the K10 was 0.88 [20]. The area under the curve signifies the probability that a randomly selected case will have a higher score on the instrument in question than a randomly chosen non-case [20]. The area under the curve was 0.96 in discriminating severe from non-severe cases of major depression [20]. In a nationally representative household survey in Australia, the K10's area under the curve was 0.90 (95% CI, 0.89 – 0.91) when differentiating 12-month cases of any mood or anxiety disorder from non-cases determined by the World Health Organization's CIDI employing DSM-IV criteria [147]. In a nationally representative household survey in Canada, the area under the curve was 0.93 (95% CI, 0.91 – 0.95) for detecting past-month and 0.87 (95% CI, 0.85 – 0.88) for past-year cases of major depressive disorder diagnosed using the WMH-CIDI [107]. The lower area under the curve for past-year cases in the Canadian sample can be attributed in part to the fact that the K10 stem question employed in that study asks about the last month rather than the worst month of the past year. In a nationally representative household survey in South Africa, when comparing the K10 inquiring about the last month with any past-year mood or anxiety disorder diagnosed using the WHO-CIDI, the overall area under the curve was 0.73, and 0.71 for major depressive disorder specifically. In Blacks (76% of the sample), the area under the curve was 0.71 for any past-year mood or anxiety disorder, compared with 0.78 in the remaining sample (comprised of those of mixed race, Whites, and Indian/Asian) (p = 0.018) [148].

Somatization. The somatization measure employed in the present study is a difference score between the somatic and psychological symptoms endorsed in the K10. Thus, the K10 is treated as a measure with two factors, somatic and psychological, in which scores are summed for each factor and the psychological sum is subtracted from the somatic sum to generate a somatization score. Despite factor analysis revealing a strong single dimension in the K10 [20], Schnittker [19] has employed this same two-factor model (referring to the psychological factor as an "affective" factor) in a previous study, assigning "tired for no good reason," "restless or fidgety," "so restless you could not sit still," and "everything an effort" to the somatic subscale, and the remaining six items to the psychological/affective subscale, a scheme used in the present study as well.

This somatization measure is modeled in two ways, as a continuous measure and a five-level categorical measure. The latter is necessitated by the fact that 543 participants in the part 2 sample (11 percent) endorsed no distress symptoms and therefore had a somatization score of 0 and contributed no information on somatization. They comprised 44 percent of all individuals with a 0 somatization score (the others with this score being those having equal scores on the somatic and psychological factors). By creating a five-level somatization variable, those with a 0 score and those at the low and high ends of somatization could be isolated, and a non-linear relationship with depression screener endorsement (expected by inclusion of participants endorsing no distress items and therefore not likely to endorse a screening symptom) could be modeled. The primary analysis for this study was done, however, among the 4,275 participants who endorsed at least one distress symptom and therefore contributed data on somatization. For this analysis, somatization was measured continuously. A secondary analysis was conducted in the full sample (i.e., including those who endorsed no distress symptoms) in which somatization was modeled both continuously and categorically.

Somatic health. These data were ascertained in the NCS-R by asking participants if a doctor or health professional had ever told them they had any of eight conditions: heart disease, high blood pressure, asthma, chronic lung disease, diabetes or high blood sugar, ulcer, epilepsy, and cancer. Participants were then asked if they still had the condition or if they were currently being treated for it; positive endorsements of this follow-up question were summed to create this measure. Scores could range from 0 to 8.

Analytic Strategy

All analyses were conducted using SAS software's (version 9.3, SAS Institute, Inc., Cary, NC) survey procedures (SURVEYMEANS, SURVEYREG, and SURVEYLOGISTIC) to account for NCS-R's survey weights and complex sampling design. Accordingly, all results reported below are weighted to the US population and the standard errors account for the multi-strata sampling design. The primary analysis was conducted in the subsample (n = 4,275) who endorsed at least one K10 item at any level and who therefore provided at least some information on somatization. A secondary analysis was conducted in the full sample (n = 4,897). All analyses adjusted for age and sex, either through standardization to the White distribution when means were compared, or in multivariable logistic regression. Differences in the means between groups for somatic health, distress, somatization, and underlying dimensions of somatization were tested using the SURVEYREG procedure, which allowed for age- and sex-standardizing using the *estimate* procedure. Significance tests of these mean differences using t-tests were conducted on *openepi.com*. The second hypothesis was tested using the SURVEYLOGISTIC procedure and adjusted for distress, age, sex, and somatization. In testing the first hypothesis in the secondary analysis, somatization was modeled both continuously and categorically. When modeled

categorically, polytomous regression was employed using the SURVEYLOGISTIC procedure with a glogit link function.

Results

This study tested a methodological explanation for the paradoxical finding from large epidemiologic studies of discordant Black-White findings on depression and distress. As a first step it was necessary to document the paradox in the NCS-R data set in the full sample. In the NCS-R, Blacks had a lower prevalence than Whites of past-year major depressive disorder, adjusting for age and sex (Black prevalence = 4.19%, White prevalence = 6.90%; t = 876.72, p < 0.0001; OR = 0.62, 95% confidence interval (CI), 0.46 − 0.83). Judging by the odds ratio and the confidence interval, the upper bound of which is well below 1.0, this was an appreciable difference. Whites had a slightly higher age- and sex-adjusted mean distress score, but the difference was not statistically significant at the p < 0.05 level (Black mean = 5.84, White mean = 6.59; t = 1.97, p = 0.05). Thus, the paradox of interest in this study was present in the NCS-R sample. Results from testing the two hypotheses are presented below, first from the primary sample and then from the secondary sample. All analyses were based on a sample weighted to reflect the US population, were age- and sex-adjusted, and drew only from past-year symptom reporting.

Primary analysis

The first hypothesis was that Blacks somatize the expression of distress more than Whites, adjusting for somatic health. The results from testing this hypothesis are reported in the first row in

Table 1, which shows that Blacks had a statistically significantly ($p < 0.05$) higher mean somatization score than Whites. Somatic health was not adjusted for in this comparison since there was no meaningful difference between Blacks and Whites on this measure (Table 1).

Table 1 also shows the results of Black-White comparisons on the underlying dimensions of the somatization measure. Among the four somatic items, Blacks had a statistically significantly higher mean score on the "everything an effort" item (this item also had the largest absolute difference in Black-White scores among the 10 items in the measure), but Whites had a statistically significantly higher mean score on the "restless or fidgety" item. Whites had higher scores than Blacks on the other two somatic items but these were not statistically significant. Among the six psychological symptoms, Whites were higher than Blacks on all, but only two of these differences ("nervous" and "hopeless") were statistically significant. In a composite measure summing mean scores across the four somatic items, Whites had a slightly higher score but the difference was not statistically significant. In a composite measure of the six psychological items, Whites were statistically significantly higher than Blacks. These results indicate that Blacks' higher somatization score was driven by their high score on one somatic item and by Whites' higher scores on the psychological items. There was no evidence of a general tendency for Blacks to endorse somatic items more than Whites; if anything, the evidence points to a general tendency for Whites to endorse psychological items at a higher level than Blacks.

The second hypothesis was that, adjusting for distress, Blacks would have a lower likelihood than Whites of endorsing a depression screening symptom, and this would be explained by Blacks' greater somatization. Results from testing this hypothesis, shown in Table 2, do not support it. Model 1 shows that Blacks had lower odds of endorsing a screening symptom, adjusting for distress, age, and sex, but the 95 percent confidence interval for race extended to 1.08. Adding somatization and somatic

health to the model (model 2) had no meaningful effect on the race odds ratio. Therefore, at the same levels of distress, Blacks did not have statistically significantly lower odds than Whites of endorsing a screening symptom, and adjusting for somatization did not alter this. Had Blacks' higher somatization constrained them from endorsing the psychological screening symptoms, adjusting for this factor in the second model would have appreciably narrowed the Black-White odds ratio of endorsing a screening symptom, or reversed it. As Table 2 shows, this did not happen.

Secondary analysis

Results from the secondary analysis, conducted in the full sample, were consistent with those from the primary analysis. As discussed above, the full sample was not used in the primary analysis because 11 percent of the full sample endorsed no past-year distress symptoms and therefore contributed no information on somatization. However, to test the robustness of the findings from the smaller, primary sample, the same hypotheses were tested in the full sample but with somatization modeled categorically in most analyses, for reasons here explained. Table S1 shows how somatization scores and frequencies were arrayed across the five somatization categories in this sample, ranging from Low (those whose actual somatization scores ranged from -16 to -3) to High (those whose actual somatization scores ranged from +2 to +12). Those with somatization scores of 0 formed the largest category and of these, 44 percent in the weighted sample endorsed no distress symptoms at all. This would induce artifactual downward pressure on the somatization-screener endorsement relationship if somatization were modeled continuously in this sample, since the modal somatization score was 0 and nearly half of these endorsed no distress symptoms and therefore had a low probability of endorsing a depression screening symptom. By modeling somatization categorically, the anticipated v-shaped relationship between somatization and screener endorsement could be specified.

Results from testing the first hypothesis – that Blacks somatize the expression of distress more than Whites – are shown in Tables S2 and S3. This hypothesis was supported when somatization was modeled continuously (for descriptive purposes) (Table S2), but not when modeled categorically (for analytic purposes) (Table S3). Thus, Table S2 shows in the first row that Blacks had a statistically significantly higher mean somatization score than Whites when somatization was modeled continuously. There was no meaningful difference on somatic health conditions between Blacks and Whites (Table S2) and so this factor was not adjusted for when comparing the groups on somatization. Black-White differences on the underlying dimensions of somatization (Table S2) closely paralleled those from the primary analysis, with no meaningful changes to highlight. This is not surprising given that the secondary sample was only 13 percent larger than the primary sample and the augmentation consisted entirely of participants who endorsed no distress items.

Table S3 compares Blacks and Whites on somatization where it was the categorical outcome of a polytomous regression, adjusting for age, sex, somatic health, and distress. Distress was adjusted for because Whites were slightly higher than Blacks on this factor and the K10 favors psychological to somatic items at a 6:4 ratio; thus higher distress scores artifactually (i.e., purely as a function of the 6:4 ratio) suppressed somatization scores. In effect, distress became a not-of-interest mediator between race and somatization, and adjusting for it removed this meditational effect. The equation modeled the odds of being in each of the four higher somatization categories compared to the odds of being in the lowest category. We see that Blacks' odds of being in higher somatization categories compared to the lowest category were not statistically significantly higher than Whites' odds, although there was a slight trend in that direction.

80

Results from testing the second hypothesis – that Blacks are less likely than Whites to endorse a screening symptom, adjusting for distress, and this is explained by Blacks' greater somatization – are shown in Table S4. They are virtually identical to the results from testing this hypothesis in the primary analysis, and therefore did not support the hypothesis. There was no statistically significant difference in Blacks' and Whites' odds of endorsing a screening symptom, both without (model 1) and with (model 2) somatization in the model.

In brief, this study found evidence of slightly greater distress somatization in Blacks than Whites, although this difference was explained by one somatic item ("everything an effort") and by Whites' general tendency to score higher than Blacks on psychological distress items. These results did not support a broad hypothesis that in the US, Blacks express distress or depression more somatically than Whites. Moreover, no evidence was found that the modest Black-White difference on somatization scores suppressed Blacks' likelihood of endorsing a depressions screening symptom.

Discussion

This study tested a methodological explanation for the consistent paradoxical finding of equal and higher distress levels in Blacks than Whites, coupled with a lower prevalence of depression. Two hypotheses were tested. The first was that Blacks express distress more somatically than Whites. The second was that Blacks' greater distress somatization explains their lower likelihood than Whites, adjusting for distress, of endorsing major depression screening symptoms. All analyses were conducted in the NCS-R data set, were adjusted for differences in the age and sex distributions between Blacks and Whites, were weighted to the US population, and were based on past-year symptom reporting.

The paradox to be explained, of a lower prevalence of depression in Blacks than Whites coupled with equal or higher levels of distress, was documented in the NCS-R sample. However, a recent systematic review [1] of Black-White distress comparisons showed higher distress levels in Blacks than Whites in 34 of 35 comparisons, though not all of these were statistically significant. In the NCS-R sample, Whites had higher distress levels than Blacks, though it was not a statistically significant difference. These distress findings mean that the paradox of interest in this sample is a slightly weaker version than might be present in other samples. Nevertheless, the relevant hypothesis in this study was that Blacks would have a lower probability than Whites of endorsing a screening symptom *adjusting for* distress, and therefore the slight divergence of the Black-White distress findings in the NCS-R sample from those of the systematic review would not have a meaningful effect on this analysis.

Although results were consistent with the first hypothesis of greater distress somatization in Blacks than Whites, the Black-White somatization difference was quite small despite being statistically significant. The meaningfulness of this small difference is further tempered by examining the underlying symptom structure of the somatization measure. The difference found on the somatization measure was driven by one somatic symptom ("everything an effort") and a greater tendency in Whites to endorse the psychological items. In fact, Whites had higher scores than Blacks on nine of the 10 items in the distress measure, though only a few of these differences were statistically significant. When somatic and psychological items were summarized in two composite scores, Whites were higher than Blacks in both, but only the psychological score difference was statistically significant. These results, if anything, provide modest support for a broad psychologization hypothesis [108] in Whites, and no support for a broad somatization hypothesis in Blacks.

Results did not support the second hypothesis. There was no statistically significant difference between Blacks' and Whites' odds of endorsing depression screening symptoms, adjusting for distress. Further adjusting for somatization did not change this.

Thus, evidence from the NCS-R does not support the idea that greater somatization in Blacks suppresses their endorsement of screening symptoms, thereby explaining the paradox of discordant Black-White comparisons on depression and distress. The evidence also does not support the notion of a generalized tendency in Blacks to express distress more somatically than Whites. Where does this leave the Black-White depression-distress paradox?

One option is to presume the depression and distress findings are valid and to propose a substantive explanation for why Blacks have less depression than Whites. To date, these proposals have included positing greater religiosity, higher self-esteem, and stronger social support [4, 36, 66–68] in Blacks than Whites – all factors thought to have dimensions protective against depression. Strong tests of the social support hypothesis have failed [66, 68] to corroborate it, however, and no empirical evidence related to the other explanations has been reported, to my knowledge. A more recent hypothesis [4, 5, 69] proposes an interaction between race, stress, and poor health behaviors (e.g., alcohol consumption) such that at higher stressor levels, these behaviors are more protective against depression in Blacks than in Whites, while simultaneously leading to worse somatic health in Blacks. Tests of the hypothesis have had mixed results [4, 69, 70].

Aside from the lack of evidence, two main problems arise with substantive explanations for the Black-White depression paradox. First, they do not account for the Black-White distress findings. That is, proponents of these explanations must explain why a given factor would protect Blacks from depression

83

but not distress, or protect Blacks from depression so much more effectively than from distress. If anything, one might think protective factors would work in the short term against the intermediate and less severe outcome of distress, but have diminishing returns over time, or in the face of larger life stressors, and ultimately fail to protect against depression. Second, the social stress paradigm predicts worse mental health outcomes in disadvantaged groups in part *by virtue* of poorer coping resources. To explain the paradox of a lower prevalence of depression in Blacks than Whites by virtue of *better* coping resources simply recreates the paradox at the locus of the hypothesized mediator, and entails a new paradox to explain.

Another option is to pursue additional artifactual explanations for the inconsistent depression and distress findings between Blacks and Whites. From the perspective of the social stress paradigm, the depression findings are the logical place to explore methodologic problems since they, rather than the distress findings, contradict the theory. One could pursue this line either as a loyal defender of the social stress paradigm or as a conservator of theory in general. That is, good or good enough theory is rare and findings contradicting it should be rigorously interrogated for bias. Alternatively, one could argue that a theory's predictions should not be expected to hold in every test of them [82], that this is too high a bar for any theory. Yet the Black-White depression comparison is a strong [15] test of the theory and results of this test should not be lightly dismissed as tolerable aberrations. That is, if the theory does not work for *this* comparison, what stock can we put in it?

Methodological explanations for the Black-White depression-distress paradox remain to be tested. As discussed in the *Introduction*, the depression interview, because of a relatively complicated algorithm that entails a series of exclusion criteria, presents more possibilities for error than the distress measures. This study considered a methodological explanation for a Black-White difference at the first

step of the algorithm, the required endorsement of a screening symptom, adjusting for distress. A similar approach could be applied to subsequent steps in the algorithm. That is, there are various points in the diagnostic interview at which individuals can be excluded from the diagnosis. The Black-White prevalence ratio of depression in a given study is determined by the proportion in each group that is excluded at each step across the interview. Thus, one could determine which step(s) in this exclusion process has the biggest impact on the final Black-White prevalence ratio in each study in which the depression paradox has been documented. A consistent pattern across studies would provide the clearest clue for where and how bias might affect Black and White depression estimates.

A recent study [149] provided evidence from one of the studies documenting the Black-White depression paradox that Blacks are substantially more likely than Whites to be excluded from the diagnosis by endorsing either the medical condition or bereavement exclusion criteria (both criteria were lumped into one question, so it is impossible to distinguish which it might have been in the relevant study, if not both). If this pattern obtains in other studies, the question is whether it is a valid reflection of the DSM criteria. If so, this can be ruled out as a source of methodological error in the Black-White depression finding. If not valid, however, and Blacks are more likely to endorse these criteria than Whites, then this suggests a point in the depression interview where methodological error accounts for at least some of the Black-White depression paradox. In short, the exclusion features of the diagnostic interview for major depression present opportunities to develop methodological explanations for the Black-White depression paradox. Any consequent explanation showing that the higher prevalence of depression in Whites owes to methodologic error in a given exclusion criterion would bring the depression findings into accordance with the distress findings and thereby resolve the second paradox.

A curious wrinkle in the Black-White depression-distress paradox is that after controlling for socio-economic status in some studies [43, 56, 86], Blacks have a lower level of distress than Whites, resulting in concordant Black-White distress and depression findings, in terms of direction. This would appear to weaken social stress theory, since controlling for socio-economic variables still leaves residual socio-economic differences and fails to account for Blacks' exposure to more interpersonal discrimination than Whites. This persistent Black-White imbalance in stressor exposure should still produce greater distress in Blacks than Whites. That it does not in these studies is consistent with the presence of a substantive factor that disproportionately protects Blacks from poor mental health outcomes. However, it could be consistent with artifactual explanations as well. Moreover, adjustment for socio-economic factors further attenuates the Black-White depression odds ratio in the NCS-R sample (results not shown), and thus the absolute value of the gap between Black-White distress and depression findings remains. Accordingly, we would still need to account for why ostensible protective factors are so much more effective against disorder than distress. But finally, the social stress paradigm generally does not parse stressor exposures into categories but rather considers the totality of stressor exposures. When this totality is considered (i.e., via Black-White comparisons adjusted at most for sex and age), Blacks generally have higher distress than Whites [1].

Strengths and limitations

The primary strengths of this study are that it used data from a large, nationally representative sample using a measure of distress that has high sensitivity for psychiatric disorder. Moreover, the depression interview and the K10 as operationalized in the NCS-R target the same time period during the past year, which should maximize concordance in their results. This high sensitivity and temporal overlap are relevant in the present study because the hypothesized reason why the depression and

distress results are discordant between Blacks and Whites is *not* that these are different constructs on which Blacks and Whites differ, but rather that these are overlapping constructs and the discordance owes to differential item response on the screening symptoms. Using a distress measure highly correlated with depression diagnoses gives us greater confidence that we are testing *this* hypothesized difference rather than differences in the constructs themselves.

Three limitations of this study are noted as well. First, depression symptoms and distress may have been inaccurately reported, leading to misclassification of the distress and screening symptom measures. To the extent this misclassification was differential by race, estimates of somatization, distress, and screening symptom endorsement may be under- or over-estimated in each race group. Second, the measure of somatic health was sub-optimal because it included only eight conditions and each required a diagnosis from a medical professional. Strong evidence exists that in the US Blacks have worse access to health care and receive poorer care than Whites [92, 134–138]. Therefore, Blacks may under-report somatic health conditions more than Whites in this study. To the extent this was true, it would inflate somatization scores to a greater extent in Blacks than Whites – regardless of whether or not somatic health is adjusted for -- and the higher level of somatization in Blacks than Whites found in this study could be an artifact. Third, in the NCS-R sample, Blacks did not have statistically significantly worse somatic health than Whites, as this variable was measured in the study. This is inconsistent with the large body of research documenting that Blacks have worse somatic health than Whites in the US [92, 139–141]. It could be due to disproportionate, though inadvertent, under-reporting of somatic conditions in Blacks, as just discussed, or to the limited number of somatic conditions measured, or to random chance in the sample.   The first two reasons have the same implications for the findings: Blacks' somatization score would be biased upward more than Whites' score. However, since there was no evidence that the small Black-White difference on somatization, as measured, suppressed Blacks'

87

endorsement of screening symptoms, this possible inflation of Blacks' somatization score would have no impact on my conclusions. On the other hand, if the somatic health of Blacks in this sample was better, by chance, than the somatic health of adult Blacks living in households nationwide, it is not clear that this would have impacted the conclusions. Black somatization scores may have been higher in a more representative Black sample, in terms of somatic health, but adjusting for somatic health ostensibly would have accounted for this.

Conclusion

In this study, although Blacks did have a higher level of distress somatization than Whites, the difference was not large and was driven by one somatic symptom and Whites' greater tendency than Blacks to psychologize distress. In the end, it was not a meaningful difference and did not constrain Blacks from endorsing the depression screening symptoms. The data from this study do not support the hypothesis that the Black-White depression-distress paradox can be explained by greater somatization in Blacks. The development and testing of additional methodological hypotheses that consider other exclusion points in the depression interview is warranted.

Table 1. Blacks' and Whites' age- and sex-standardized means and prevalences among those endorsing at least one distress symptom in the National Comorbidity Survey Replication (2001-2003)

	Black 611 (14.13%)[1]	White 3,664 (85.87%)	t value	p value
Somatization mean (SE), range -24 to +16	-0.02 (0.15)	-0.36 (0.05)	-2.48	0.01
Somatic health conditions mean (SE), range 0-7	1.03 (0.06)	1.01 (0.04)	-0.20	0.84
Distress mean (SE), range 0 - 40	7.14 (0.36)	7.72 (0.14)	1.56	0.12
Underlying dimensions of somatization and depression				
Somatic symptom means (SE), range 0-4				
Low energy	1.19 (0.07)	1.28 (0.02)	1.59	0.11
Restless or fidgety	0.83 (0.06)	1.05 (0.02)	4.02	< 0.01
Couldn't sit still	0.40 (0.04)	0.48 (0.02)	1.55	0.12
Everything an effort	1.15 (0.09)	0.87 (0.03)	-3.41	< 0.01
Psychological symptom means (SE), range 0-4				
Nervous	1.06 (0.06)	1.21 (0.02)	2.74	0.01
Couldn't calm down	0.31 (0.04)	0.35 (0.02)	0.78	0.44
Hopeless	0.42 (0.05)	0.55 (0.02)	2.45	0.01
Depressed	1.01 (0.05)	1.07 (0.02)	1.13	0.26
Couldn't cheer up	0.43 (0.05)	0.45 (0.02)	0.38	0.71
Worthless	0.35 (0.03)	0.42 (0.02)	1.39	0.17
Somatic symptoms, mean (SE), range 0-16	3.56 (0.18)	3.68 (0.07)	0.64	0.52
Psychological symptoms, mean (SE), range 0-24	3.58 (0.21)	4.04 (0.08)	2.15	0.03

Note. All comparisons are based on weighted sample.

[1] unweighted *n* (weighted %)

Table 2. Logistic regression of race predicting screener endorsement, adjusting for distress, age, sex, somatization, and somatic health among those endorsing at least one distress symptom in the National Comorbidity Survey Replication (2001-2003)

| | Model 1 | | Model 2 | |
	OR	95% CI	OR	95% CI
Race (ref = White)	0.75	0.52 - 1.08	0.77	0.54 - 1.10
Distress	1.18	1.16 - 1.19	1.13	1.11 - 1.15
Age	0.84	0.76 - 0.93	0.78	0.71 - 0.87
Sex (ref = male)	1.11	0.91 - 1.37	1.03	0.84 - 1.26
Somatization			0.88	0.84 - 0.91
Somatic health			1.34	1.21 - 1.49

Note. Both analyses used weighted sample.

**Table S1. Somatization categories in the full sample,
National Comorbidity Survey Replication (2001-2003)**

Somatization category	Actual somatization scores	n (%)[1]
Low	-16 to -3	965 (14.7)
Medium low	-2 and -1	1,014 (20.8)
Medium	0	1,223 (29.8)
Medium high	1 and 2	1,119 (24.6)
High	3 to 12	497 (10.2)

[1] unweighted n (weighted %)

Table S2. Blacks' and Whites' age-and sex-standardized means and prevalences in the full sample, National Comorbidity Survey Replication (2001-2003)

	Black	White		
	717 (14.53%)[1]	4,180 (85.47%)	*t* value	*p* value
Somatization mean (SE), range -24 to +16	-0.03 (0.13)	-0.33 (0.04)	-2.71	0.01
Somatic health conditions mean (SE), range 0-7	0.92 (0.05)	0.97 (0.04)	0.51	0.61
Distress mean (SE), range 0 - 40	5.84 (0.33)	6.59 (0.15)	1.97	0.05
Underlying dimensions of somatization and depression				
Somatic symptom means (SE), range 0-4				
Low energy	0.95 (0.06)	1.08 (0.03)	1.70	0.09
Restless or fidgety	0.66 (0.05)	0.89 (0.02)	4.38	<.0001
Couldn't sit still	0.31 (0.03)	0.41 (0.02)	2.01	0.05
Everything an effort	0.92 (0.07)	0.74 (0.02)	-3.20	0.001
Psychological symptom means (SE), range 0-4				
Nervous	0.87 (0.05)	1.03 (0.02)	3.05	0.002
Couldn't calm down	0.26 (0.03)	0.30 (0.02)	0.80	0.42
Hopeless	0.34 (0.04)	0.47 (0.02)	2.55	0.01
Depressed	0.82 (0.05)	0.91 (0.02)	1.71	0.09
Couldn't cheer up	0.35 (0.05)	0.39 (0.01)	1.26	0.21
Worthless	0.28 (0.03)	0.35 (0.01)	2.58	0.01
Somatic symptoms, mean (SE), range 0-16	2.84 (0.17)	3.11 (0.08)	1.31	0.19
Psychological symptoms, mean (SE), range 0-24	2.94 (0.19)	3.47 (0.07)	2.84	0.01

Note. All comparisons are based on weighted sample.

[1] unweighted *n* (weighted percent)

Table S3. Polytomous regression of somatization on race, adjusting for age, sex, somatic health, and distress in the full sample, National Comorbidity Survey Replication (2001-2003)

	Somatization (ref = low)[1]	OR	95% CI
Race (ref = White)	medium low	1.00	0.60 - 1.69
	medium	1.00	0.66 - 1.50
	medium high	1.08	0.64 - 1.83
	high	1.25	0.70 - 2.22
Age	medium low	0.96	0.80 - 1.15
	medium	1.11	0.95 - 1.30
	medium high	0.94	0.79 - 1.13
	high	1.15	0.94 - 1.40
Sex	medium low	0.89	0.68 - 1.17
	medium	0.89	0.67 - 1.19
	medium high	0.93	0.71 - 1.22
	high	0.64	0.48 - 0.85
Somatic health	medium low	1.04	0.94 - 1.15
	medium	1.00	0.86 - 1.15
	medium high	1.14	1.00 - 1.29
	high	1.20	1.06 - 1.37
Distress	medium low	0.84	0.82 - 0.85
	medium	0.67	0.64 - 0.70
	medium high	0.78	0.77 - 0.80
	high	0.87	0.85 - 0.88

Note. Analysis used weighted sample.
[1] Actual somatization scores in each category are: "low" (ref) (-16 to -3); "medium low" (-2 and -1) "medium" (0, which includes those endorsing no symptoms); "medium high" (+1 and +2); and "high" (+3 to +12).

Chapter 5: CONCLUSION

This dissertation has roots in my interest in American race history and my first glimpse nine years ago of the Black-White depression finding. Before knowing anything about social stress theory, I found it richly ironic that arguably the most marginalized group in American history turned out to have less depression than the marginalizing group. A sample of one is not exactly epidemiology, but my reaction suggests that the core principal of social stress theory – social disadvantage is bad for mental health -- is not an academic abstraction but perhaps a common or even universal intuition.

Nine years later, my motivation to search for a methodological explanation for the double paradox stems from both disenchantment with substantive explanations and an interest in taking a conservative approach to doing science. As noted throughout my dissertation, the substantive explanations proffered to date do not come to terms with the distress findings and therefore fail to be conceptually convincing. Regarding a conservative approach to doing science, systematically testing methodological explanations for the counter-intuitive double paradox seems more efficient and logical than prematurely rejecting a theory, or proceeding a-theoretically. Consequently, methodological explanations became my focus.

Results from the first paper make it clear that the double paradox does not reflect a selective reading of the literature but is an extremely consistent finding. The somatization hypothesis offered one solution to both paradoxes. It would account for both a lower prevalence of depression in Blacks than Whites and for why this pattern does not occur with distress. It also offered a chance to conduct a good test of an idea frequently found in the literature – greater depression somatization in Blacks than

94

Whites -- and to either find support for it or lay it to rest. This dissertation found no support for a broad hypothesis of greater somatization in Blacks than Whites. Instead what it found were two different somatic symptoms – one in each study – which Blacks were much more likely to endorse than Whites and a broad tendency in Whites to endorse psychological symptoms more than Blacks. However, these two somatic symptoms and the tendency to endorse psychological symptoms in Whites was not sufficient to create either a big difference between Blacks and Whites on the somatization measure, or for that small difference to explain Black and White screener endorsement patterns. Thus, this dissertation's findings ultimately do not support a meaningful difference between Blacks and Whites in their symptom endorsement patterns.

This conclusion of no consequential Black-White difference on somatization is based on data from two of the four nationally representative psychiatric epidemiology studies that document the double paradox. The two other studies (the National Comorbidity Survey, and the National Epidemiologic Survey on Alcohol and Related Conditions) do not lend themselves to testing the somatization hypothesis because they lack measures in which somatization can be adequately assessed. Nevertheless, this dissertation's tests of the somatization hypothesis, though far from perfect because the data were not collected with this question in mind, are good enough to make me inclined to rule this explanation out in the future. They are good enough tests of somatization because the AUDADIS-IV depression interview and the K10 were employed in representative samples of the US household population and both have a good balance of the common somatic and psychological expressions of depression and distress, respectively. As well, the results were consistent across the two papers and reliability is a necessary condition of validity.

The diagnostic interview for major depression winnows individuals from a depression diagnosis across seven exclusion criteria: not endorsing a screening symptom; not endorsing five or more symptoms; not being clinically significant; ever having had a manic episode; and, the depression being better accounted for by bereavement within two months of a loss, by a medical condition physiologically causing the depression, or by use of or withdrawal from alcohol or a medication or drug. The key then to understanding why Blacks have a lower prevalence than Whites of major depression in the psychiatric epidemiology studies is to see which of these seven steps accounts for the greatest attenuation in the Black-White prevalence ratio across the interview. Not all of the psychiatric epidemiology studies operationalize all of these steps, and there is variability in how well they implement them, so this approach is best leveraged by studying these patterns across studies to see where the preponderance of the evidence lies.

The subsequent step would be to understand why the attenuation in the Black-White prevalence ratio happens where it does, taking into account both methodological and substantive explanations. Distinguishing between these two categories of explanation in this context is not necessarily easy. For example, if the greatest attenuation occurred with the clinical significance criterion, one might look for consistency with other information study participants provide in the full study interview to help determine whether the attenuation was due to artifact or a substantive reason. Participants' biased reporting is suggested when this other information is inconsistent with endorsing, or failing to endorse, clinical significance. Inconsistency that is more common in Blacks or Whites points to a methodological explanation for the Black-White depression paradox. Alternatively, no Black-White differences on inconsistent reporting points to a substantive explanation for why Blacks are more likely to be excluded from the diagnosis at this stage in the diagnostic interview.

The question then is what next steps should be taken towards resolving the double paradox. One clear approach is suggested in results from the second paper, where it was shown that Blacks were nearly 50 percent more likely than Whites to say that their past-year depression symptoms co-occurred with physical illness or bereavement and who were consequently excluded from the diagnosis. However, this was an overly broad operationalization of the DSM exclusion criteria that most certainly led to invalid exclusions that were most likely disproportionately Black. Future research should examine how carefully these exclusion criteria were operationalized in other studies documenting the Black-White depression paradox, and if meaningful differences in Black and White endorsement patterns persist with more precise rendering of these criteria. If this pattern does persist across more careful renderings of the criteria, it becomes less plausible that these specific criteria contribute to bias in Black-White depression prevalence ratio estimates. At stake is that greater morbidity and mortality in Blacks than Whites could lead to disproportionately greater invalid exclusions in Blacks than Whites based on poorly operationalized medical condition and bereavement exclusions, as was most likely demonstrated in the NLAES data set. Blacks' greater morbidity and mortality than Whites are legitimate causes of major depression outside of cases directly and physiologically caused by a medical condition, and inappropriate winnowing of these legitimate cases from a diagnosis could explain much of the Black-White depression paradox.

As noted throughout the dissertation, it is incumbent on all substantive explanations for the Black-White depression paradox to also account for the contrasting Black-White distress findings. Not because depression and distress are synonymous but because they are similar enough that whatever protects against depression ought to also protect against distress. Moreover, social stress theory makes similar predictions for disorder and distress.

It is also important to note that the development of methodological and substantive explanations should focus on Whites as much as Blacks. Accordingly, it should not be assumed that estimates of disorder and distress in Whites are less biased than they might be in Blacks. For example, Whites may be overly prone to endorse diagnostic symptoms because of anti-depressant media campaigns targeting them or because of cultural factors in which problems are too readily reduced to or described by psychological factors or language. On the substantive side, harmful factors more prevalent in Whites are just as likely to occur as protective factors in Blacks. For instance, cultural atomization may play an etiologic role in psychiatric disorder and may also be more widespread among Whites than Blacks [150–152]. As well, the same substantive explanation could be framed from either the White or Black perspective: whatever harmful factor is posited to be more prevalent in Whites, its opposite could be posited to be more prevalent in Blacks. Thus atomism more prevalent in Whites could be re-framed as communalism more prevalent in Blacks. The bottom line is that findings in Whites should never be viewed as inerrant or as necessarily defining acceptable norms.

A final, broader question, and one implicit throughout this dissertation, is what the implications of my findings are for the social stress paradigm. I framed Black-White mental health comparisons as a strong test of this paradigm, and the Black-White depression paradox as a potential threat to the paradigm's credibility. Having ruled out a plausible methodological explanation for the double paradox, how do I now view the social stress paradigm in the context of a still unresolved double paradox, while also taking into account that substantive explanations have performed no better than methodological explanations at resolving the paradoxes? In the context of the still unresolved double paradox, and in light of the fact that methodological explanations remain to be tested, the social stress paradigm remains viable. However, the paradigm must be judged in a broader context. After all, this is not the only epidemiologic paradox from the vantage point of social stress theory. Findings on mental health

outcomes by sex [153] and social class [32, 84] also fail to conform neatly to the theory's predictions since women and men have approximately equal levels of mental and substance use disorders, and the findings for social class are uneven depending on how one measures class (e.g., education, income), and even within a given measure the findings vary across studies. Sexual orientation is one exception to this pattern, given that lesbians, gays, and bisexuals are generally found to have a higher prevalence of psychiatric disorder than heterosexuals [154, 155]. Despite this exception, across four central axes of social division (race, sex, class, and sexual orientation), the empirical evidence supporting social stress theory is not strong. This broader context of uneven findings could make one less sanguine about social stress theory's viability.

However, is the problem with the social stress paradigm or with how we measure mental health? Intriguingly, the distress findings across the four social axes defined in the previous paragraph do conform better to social stress paradigm predictions than do the psychiatric disorder findings [1, 156–158]. Therefore, if there is a measurement problem, then perhaps, as proposed in this dissertation, the greater problem is in how we measure disorder, not how we measure distress. The evidence for criterion validity in the psychiatric epidemiology studies, measured as diagnostic or prevalence concordance between structured lay interview and clinician interview diagnoses is poor [32, 67, 159]. The National Institute of Mental Health's Research Domain Criteria project (RDoC), which seeks to redefine mental disorder according to objective physiological markers rather than symptom self-report to improve the validity of psychiatric disorder diagnoses, underscores this apparent weakness in current measures of mental disorder. It is arguably then too optimistic to presume general validity in the psychiatric epidemiology studies regarding group differences in disorder. Until the validity of psychiatric diagnosis improves, it would be precipitous to count social stress theory out on the basis of dis-confirming findings from psychiatric epidemiology. In the meantime, a focus on methodological

explanations for unexpected findings in psychiatric epidemiology – from the viewpoint of the social

stress paradigm – is a prudent path to take.

References

1. Barnes DM (manuscript in preparation) Documenting a double paradox: a systematic literature review comparing Blacks and Whites in the US on major depression and psychological distress.

2. Brown TN (2003) Critical race theory speaks to the sociology of mental health: mental health problems produced by racial stratification. J Health Soc Behav 44:292–301.

3. Earl T, Williams D, Anglade S (2011) An Update on the Mental Health of Black Americans: Puzzling Dilemmas and Needed Research. J Black Psychol 37:485–498. doi: 10.1177/0095798410396077

4. Jackson JS, Knight KM, Rafferty JA (2010) Race and unhealthy behaviors: chronic stress, the HPA axis, and physical and mental health disparities over the life course. Am J Public Health 100:933–939.

5. Mezuk B, Abdou CM, Hudson D, et al. (2013) "White Box" Epidemiology and the Social Neuroscience of Health Behaviors: The Environmental Affordances Model. Soc Ment Health 3:79–95. doi: 10.1177/2156869313480892

6. Neighbors HW, Njai R, Jackson JS (2007) Race, ethnicity, John Henryism, and depressive symptoms: The national survey of American life adult reinterview. Res Hum Dev 4:71–87.

7. Dressler WW, Oths KS, Gravlee CC (2005) Race and ethnicity in public health research: models to explain health disparities. Annu Rev Anthr 34:231–252.

8. Horwitz AV (2002) Outcomes in the Sociology of Mental Health: Where Have We Been and Where are We Going? J Health Soc Behav 43:143–151.

9. Pearlin LI, Meneghan EG, Lieberman MA, Mullan JT (1981) The Stress Process. J Health Soc Behav 22:337–356.

10. Schwartz S, Meyer IH (2010) Mental health disparities research: the impact of within and between group analyses on tests of social stress hypotheses. Soc Sci Med 70:1111–1118.

11. Turner RJ (2010) Health disparities and the stress hypothesis: A commentary on Schwartz and Meyer. Soc Sci Med 70:1119–1120.

12. Muntaner C, Ng E, Vanroelen C, et al. (2013) Social Stratification, Social Closure, and Social Class as Determinants of Mental Health Disparities. In: Aneshensel CS, Phelan JC, Bierman A (eds) Handb. Sociol. Ment. Health. Springer Netherlands, Dordrecht, pp 205–227

13. Jordan WD (1968) White Over Black: American Attitudes Toward the Negro, 1550-1812. The University of North Carolina Press, Chapel Hill

14. Williams DR, Yu Y, Jackson JS, Anderson NB (1997) Racial differences in physical and mental health, socio-economic status, stress and discrimination. J Health Psychol 2:335–351.

15. Platt JR (1964) Strong inference. Science 146:347–353.

16. Lincoln KD, Taylor RJ, Watkins DC, Chatters LM (2010) Correlates of Psychological Distress and Major Depressive Disorder Among African American Men. Res Soc Work Pract 21:278–288. doi: 10.1177/1049731510386122

17. Somervell PD, Leaf PJ, Weissman MM, et al. (1989) The prevalence of major depression in black and white adults in five United States communities. Am J Epidemiol 130:725–735.

18. Dohrenwend BP, Shrout PE, Egri G, Mendelsohn FS (1980) Nonspecific psychological distress and other dimensions of psychopathology: Measures for use in the general population. Arch Gen Psychiatry 37:1229.

19. Schnittker J (2012) The Proximity of Common Unhappiness and Misery. Soc Ment Health 2:135–153. doi: 10.1177/2156869312451153

20. Kessler RC, Andrews G, Colpe LJ, et al. (2002) Short screening scales to monitor population prevalences and trends in non-specific pscyhological distress. Psychol Med 32:959–976.

21. Link B, Dohrenwend BP (1980) Formulation of Hypotheses About the True Prevalence of Demoralization in the United States. Ment. Illn. U. S. Epidemiol. Estim. Praeger Publishers, New York, pp 114–132

22. Coyne JC (1994) Self-reported distress: Analog or ersatz depression? Psychol Bull 116:29–45.

23. Horwitz AV, Wakefield JC (2007) The Loss of Sadness: How Psychiatry Transformed Normal Sorrow Into Depressive Disorder. Oxford University Press, New York

24. Andrews G, Slade T (2001) Interpreting scores on the Kessler psychological distress scale (K10). Aust N Z J Public Health 25:494–497.

25. Fechner-Bates S, Coyne JC, Schwenk TL (1994) The relationship of self-reported distress to depressive disorders and other psychopathology. J Consult Clin Psychol 62:550.

26. Kroenke K, Spitzer RL, Williams JB (2001) The Phq-9. J Gen Intern Med 16:606–613.

27. Breslau J, Javaras KN, Blacker D, et al. (2008) Differential item functioning between ethnic groups in the epidemiological assessment of depression. J Nerv Ment Dis 196:297–306.

28. Dohrenwend BP (2000) The role of adversity and stress in psychopathology: some evidence and its implications for theory and research. J Health Soc Behav 41:1–19.

29. Monroe SM, Slavich GM, Georgiades K (2009) The social environment and life stress in depression. Handb. Depress., 2nd ed. The Guilford Press, New York, pp 340–360

30. Shrout PE, Link BG, Dohrenwend BP, et al. (1989) Characterizing life events as risk factors for depression: the role of fateful loss events. J Abnorm Psychol 98:460.

31. Turner RJ, Lloyd DA (2004) Stress burden and the lifetime incidence of psychiatric disorder in young adults: racial and ethnic contrasts. Arch Gen Psychiatry 61:481.

32. Kessler RC, Berglund P, Demler O, et al. (2005) Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 62:593.

33. Smith SM, Stinson FS, Dawson DA, et al. (2006) Race/ethnic differences in the prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. Psychol Med 36:987. doi: 10.1017/S0033291706007690

34. Kroenke K, Strine TW, Spitzer RL, et al. (2009) The PHQ-8 as a measure of current depression in the general population. J Affect Disord 114:163–173. doi: 10.1016/j.jad.2008.06.026

35. Blazer DG, Kessler RC (1994) The prevalence and distribution of major depression in a national community sample: the National Comorbidity Survey. Am J Psychiatry 151:979–986.

36. Breslau J, Aguilar-Gaxiola, S., Kendler KS, et al. (2006) Specifying race-ethnic differences in risk for psychiatric disorder in a US national sample. Psychol Med 36:57–68.

37. Compton W, Conway K, Stinson F, Grant B (2006) Changes in the prevalence of major depression and comorbid substance use disorders in the United States between 1991–1992 and 2001–2002. Am J Psychiatry 163:2141–2147.

38. Hasin DS, Goodwin RD, Stinson FS, Grant BF (2005) Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. Arch Gen Psychiatry 62:1097–1106.

39. Regier DA, Farmer ME, Rae DS, et al. (1993) One-month prevalence of mental disorders in the United States and sociodemographic characteristics: the Epidemiologic Catchment Area study. Acta Psychiatr Scand 88:35–47.

40. Weissman MM, Bruce ML, Leaf PJ, et al. (1991) Affective disorders. Psychiatr. Disord. Am. Simon and Schuster, New York, pp 53–80

41. Satcher D (2001) Mental health: Culture, race, and ethnicity—A supplement to mental health: A report of the surgeon general.

42. Bratter JL, Eschbach K (2005) Race/Ethnic Differences in Nonspecific Psychological Distress: Evidence from the National Health Interview Survey*. Soc Sci Q 86:620–644.

43. Centers for Disease and Control and Prevention (2004) Self-reported frequent mental distress among adults - United States, 1993-2001. 963–966.

44. Centers for Disease and Control and Prevention (2010) Current depression among adults -- United States, 2006 and 2008. 1229–1235.

45. Chowdhury PP, Balluz L, Strine TW (2008) Health-related quality of life among minority populations in the United States, BRFSS 2001-2002. Ethn Dis 18:483–487.

46. Dey AN, Lucas JW, others (2006) Physical and mental health characteristics of US-and foreign-born adults, United States, 1998-2003. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics

47. Dismuke CE, Egede LE (2011) Association of serious psychological distress with health services expenditures and utilization in a national sample of US adults. Gen Hosp Psychiatry 33:311–317. doi: 10.1016/j.genhosppsych.2011.03.014

48. Eaton W, Kessler LG (1981) Rates of symptoms of depression in a national sample. Am J Epidemiol 114:528–538.

49. Fiscella K, Franks P (1997) Does psychological distress contribute to racial and socio-economic disparities in mortality? Soc Sci Med 45:1805–1809.

50. Fleischman JA (2007) Measurement equivalence for three mental health status measures. Department of Health and Human Services, Peachtree City, GA, pp 55–62

51. Harris KM, Edlund MJ, Larson S (2005) Racial and ethnic differences in the mental health problems and use of mental health care. Med Care 43:775–784.

52. Jones-Webb RJ, Snowden LR (1993) Symptoms of depression among Blacks and Whites. Am J Public Health 83:240–244.

53. Kessler RC, Neighbors HW (1986) A new perspective on the relationships among race, social class, and psychological distress. J Health Soc Behav 27:107–115.

54. Kiviniemi MT, Orom H, Giovino GA (2011) Psychological Distress and Smoking Behavior: The Nature of the Relation Differs by Race/Ethnicity. Nicotine Tob Res 13:113–119. doi: 10.1093/ntr/ntq218

55. Mellinger GD, Balter MB, Manheimer DI, et al. (1978) Psychic distress, life crisis, and use of psychotherapeutic medications: national household survey data. Arch Gen Psychiatry 35:1045–1052.

56. Mukherjee S (2013) Comparing Adult Males and Females in the United States to Examine the Association between Body Mass Index and Frequent Mental Distress: An Analysis of Data from BRFSS 2011. Psychiatry J 2013:1–11. doi: 10.1155/2013/230928

57. Mulia N, Ye Y, Zemore SE, Greenfield TK (2008) Social disadvantage, stress, and alcohol use among black, Hispanic, and white Americans: findings from the 2005 US National Alcohol Survey. J Stud Alcohol Drugs 69:824.

58. Nuru-Jeter A, Williams CT, LaVeist TA (2008) A methodological note on modeling the effects of race: the case of psychological distress. Stress Health 24:337–350. doi: 10.1002/smi.1215

59. Oraka E (2010) Asthma and Serious Psychological Distress: Prevalence and Risk Factors Among US Adults, 2001-2007. CHEST J 137:609. doi: 10.1378/chest.09-1777

60. Pratt LA (2009) Serious Psychological Distress, as Measured by the K6, and Mortality. Ann Epidemiol 19:202–209. doi: 10.1016/j.annepidem.2008.12.005

61. Reeves WC, Strine TW, Pratt LA, et al. (2011) Mental illness surveillance among adults in the United States. 1–29.

62. Roxburgh S (2009) Untangling Inequalities: Gender, Race, and Socioeconomic Differences in Depression. Sociol Forum 24:357–381. doi: 10.1111/j.1573-7861.2009.01103.x

63. Strine TW, Dhingra SS, Okoro CA, et al. (2009) State-based differences in the prevalence and characteristics of untreated persons with serious psychological distress. Int J Public Health 54:9–15. doi: 10.1007/s00038-009-0001-6

64. Wu L-T, Anthony JC (2000) The estimated rate of depressed mood in US adults: recent evidence for a peak in later life. J Affect Disord 60:159–171.

65. Zahran HS, Kobau R, Moriarty DG, et al. (2005) Health-related quality of life surveillance—United States, 1993–2002. 1–35.

66. Mouzon DM (2010) Can social relationships explain the race paradox in mental health? Rutgers University-Graduate School-New Brunswick

67. Williams DR, Gonzalez HM, Neighbors H, et al. (2007) Prevalence and distribution of major depressive disorder in African Americans, Caribbean blacks, and non-Hispanic whites: results from the National Survey of American Life. Arch Gen Psychiatry 64:305.

68. Kiecolt K, Hughes M, Keith VM (2008) Race, social relationships, and mental health. Pers Relatsh 15:229–245.

69. Mezuk B, Rafferty JA, Kershaw KN, et al. (2010) Reconsidering the role of social disadvantage in physical and mental health: stressful life events, health behaviors, race, and depression. Am J Epidemiol 172:1238–1249.

70. Keyes KM, Barnes DM, Bates LM (2011) Stress, coping, and depression: testing a new hypothesis in a prospectively studied general population sample of U.S.-born whites and blacks. Soc Sci Med 72:650–659.

71. Keyes CLM (2009) The black–white paradox in health: flourishing in the face of social inequality and discrimination. J Pers 77:1677–1706.

72. Mays VM, Cochran SD, Ponce NA (2004) Thinking about race and ethnicity in population-based studies of health. In: Beech BM, Goodman M (eds) Race Res. Perspect. Minor. Particip. Health Stud. American Public Health Association, Washington, D.C., pp 79–100

73. Williams DR, Harris-Reid M (1999) Race and mental health: emerging patterns and promising approaches. In: Horwitz AV, Scheid TL (eds) Handb. Study Ment. Health Soc. Contexts Theor. Syst. Cambridge University Press, Cambridge, pp 295–314

74. Barnes DM, Keyes KM, Bates LM (2013) Racial differences in depression in the United States: how do subgroup analyses inform a paradox? Soc Psychiatry Psychiatr Epidemiol 48:1941–1949. doi: 10.1007/s00127-013-0718-7

75. Alegria M, McGuire T (2003) Rethinking a universal framework in the psychiatric symptom-disorder relationship. J Health Soc Behav 44:257–274.

76. Uebelacker LA, Strong D, Weinstock LM, Miller IW (2008) Use of item response theory to understand differential functioning of DSM-IV major depression symptoms by race, ethnicity and gender. Psychol Med 39:591. doi: 10.1017/S0033291708003875

77. Ayalon L, Young MA (2003) A Comparison Of Depressive Symptoms In African Americans And Caucasian Americans. J Cross-Cult Psychol 34:111–124. doi: 10.1177/0022022102239158

78. Baker FM (2001) Diagnosing depression in African Americans. Community Ment Health J 37:31–38.

79. Carter JH (1974) Recognizing psychiatric symptoms in black Americans. Geriatrics 95–99.

80. Das AK, Olfson M, McCurtis HL, Weissman MM (2006) Depression in African Americans: breaking barriers to detection and treatment. J Fam Pract 55:30.

81. Fabrega Jr H, Mezzich J, Ulrich RF (1988) Black-white differences in psychopathology in an urban psychiatric population. Compr Psychiatry 29:285–297.

82. Aneshensel CS (1992) Social Stress: Theory and Research. Annu Rev Sociol 18:15–38.

83. Hasin DS, Stinson FS, Ogburn E, Grant BF (2007) Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Arch Gen Psychiatry 64:830–842.

84. Kessler RC, McGonagle KA, Zhao S, et al. (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. Arch Gen Psychiatry 51:8–19.

85. Hanna EZ, Grant BF (1999) NLAES smoking drugs MDD Hanna & Grant 1999.pdf. Alcohol Clin Exp Res 23:513–522.

86. Strine TW, Dhingra SS, Okoro CA, et al. (2009) State-based differences in the prevalence and characteristics of untreated persons with serious psychological distress. Int J Public Health 54:9–15. doi: 10.1007/s00038-009-0001-6

87. Aneshensel CS (1999) Outcomes of the Stress Process. Handb. Study Ment. Health Soc. Contexts Theor. Syst. Cambridge University Press, Cambridge, pp 211–227

88. Aneshensel CS, Phelan JC (1999) The Sociology of Mental Health: Surveying the Field. Handb. Sociol. Ment. Health. Plenum Publishers, New York, pp 3–18

89. Pearlin LI, Bierman A (2013) Current Issues and Future Directions in Research into the Stress Process. In: Aneshensel CS, Phelan JC, Bierman A (eds) Handb. Sociol. Ment. Health. Springer Netherlands, Dordrecht, pp 325–340

90. Thoits PA (1999) Sociological Approaches to Mental Health. Handb. Study Ment. Health Soc. Contexts Theor. Syst. Cambridge University Press, Cambridge, pp 121–138

91. Wheaton B (1999) The Nature of Stressors. Handb. Study Ment. Health Soc. Contexts Theor. Syst. Cambridge University Press, Cambridge, pp 176–197

92. Williams DR, Jackson PB (2005) Social Sources Of Racial Disparities In Health. Health Aff (Millwood) 24:325–334. doi: 10.1377/hlthaff.24.2.325

93. Kessler RC (2002) The categorical versus dimensional assessment controversy in the sociology of mental illness. J Health Soc Behav 43:171–188.

94. Mirowsky J, Ross CE (2002) Measurement for a Human Science. J Health Soc Behav 43:152–170.

95. Bertera EM (2005) Mental health in U.S. adults: The role of positive social support and social negativity in personal relationships. J Soc Pers Relatsh 22:33–48. doi: 10.1177/0265407505049320

96. Chatters LM, Taylor RJ, Jackson JS, Lincoln KD (2008) Religious coping among African Americans, Caribbean Blacks and Non-Hispanic Whites. J Community Psychol 36:371–386. doi: 10.1002/jcop.20202

97. Ellison CG, Boardman JD, Williams DR, Jackson JS (2001) Religious involvement, stress, and mental health: Findings from the 1995 Detroit Area Study. Soc Forces 80:215–249.

98. Fischer AR, Shaw CM (1999) African Americans' mental health and perceptions of racist discrimination: The moderating effects of racial socialization experiences and self-esteem. J Couns Psychol 46:395.

99. Hettler TR, Cohen LH (1998) Intrinsic religiousness as a stress-moderator for adult Protestant churchgoers. J Community Psychol 26:597–609.

100. Gray BA, Keith VM (2003) The benefits and costs of social support for African American women. Our Right Minds Ment. Health Afr. Am. Women. Columbia University Press, New York, pp 242–257

101. Hughes D, Rodriguez J, Smith EP, et al. (2006) Parents' ethnic-racial socialization practices: A review of research and directions for future study. Dev Psychol 42:747–770. doi: 10.1037/0012-1649.42.5.747

102. Van Olphen J, Schulz A, Israel B, et al. (2003) Religious involvement, social support, and health among African-American women on the east side of Detroit. J Gen Intern Med 18:549–557.

103. Jackson JS, Knight KM (2006) Race and self-regulatory health behaviors: the role of the stress response and the HPA axis in physical and mental health disparities. In: Schaie KW, Carstensen LL (eds) Soc. Struct. Aging Self-Regul. Elder. Springer, New York, pp 189–207

104. Neighbors HW, Trierweiler SJ, Munday C, et al. (1999) Psychiatric diagnosis of African Americans: diagnostic divergence in clinician-structured and semistructured interviewing conditions. J Natl Med Assoc 91:601–612.

105. Neighbors HW, Trierweiler SJ, Ford BC, Muroff JR (2003) Racial differences in DSM diagnosis using a semi-structured instrument: the importance of clinical judgment in the diagnosis of African Americans. J Health Soc Behav 237–256.

106. Strakowski SM, Keck PE, Arnold LM, et al. (2003) Ethnicity and diagnoses in patients with affective disorders. J Clin Psychiatry 64:747–754.

107. Cairney J, Veldhuizen BA, Wade TJ, et al. (2007) Evaluation of 2 measures of psychological distress as screeners for depression in the general population. Can J Psychiatry 52:111–120.

108. Simon GE, VonKorff M, Piccinelli M, et al. (1999) An international study of the relation between somatic symptoms and depression. N Engl J Med 341:1329–1335.

109. Lehmann HE (1959) Psychiatric concepts of depression: nomenclature and classification. Can Psychiatr Assoc J 4:1–12.

110. Cadoret RJ, Wilson D (1983) Somatization among depressed patients in industrialized nations. Am J Psychiatry 140:1103–1104.

111. Carragher N, Adamson G, Bunting B, McCann S (2009) Subtypes of depression in a nationally representative sample. J Affect Disord 113:88–99. doi: 10.1016/j.jad.2008.05.015

112. Kirmayer LJ, Robbins JM, Dworkind M, Yaffe MJ (1993) Somatization and the recognition of depression and anxiety in primary care. Am J Pyschiatry 150:734–741.

113. Kirmayer LJ, Young A (1998) Culture and somatization: clinical, epidemiological, and ethnographic perspectives. Psychosom Med 60:420–430.

114. Kleinman AM (1977) Depression, somatization and the "new cross-cultural psychiatry."Soc Sci Med 11:3–10.

115. Ryder AG, Yang J, Heine SJ (2002) Somatization vs. psychologization of emotional distress: A paradigmatic example for cultural psychopathology. Online Read Psychol Cult 10:3.

116. Tseng W-S, Hsu J (1970) Chinese Culture, Personality Formation and Mental Illness. Int J Soc Psychiatry 16:5–14. doi: 10.1177/002076407001600101

117. Tylee A, Gandhi P (2005) The importance of somatic symptoms in depression in primary care. Prim Care Companion J Clin Psychiatry 7:167–176.

118. Yusim A, Anbarasan D, Hall B, et al. (2010) Sociocultural domains of depression among indigenous populations in Latin America. Int Rev Psychiatry 22:370–377. doi: 10.3109/09540261.2010.500870

119. Ying Y-W, Lee PA, Tsai JL, et al. (2000) The conception of depression in Chinese American college students. Cultur Divers Ethnic Minor Psychol 6:183–195. doi: 10.1037/1099-9809.6.2.183

120. Kirmayer LJ, Looper KJ (2006) Abnormal illness behaviour: physiological, psychological and social dimensions of coping with distress. Curr Opin Psychiatry 19:54–60.

121. Parker G, Cheah Y-C, Roy K (2001) Do the Chinese somatize depression? A cross-cultural study. Soc Psychiatry Psychiatr Epidemiol 36:287–293.

122. Prange AJ, Vitols MM (1962) Cultural Aspects of the Relatively Low Incidence of Depression in Southern Negroes. Int J Soc Psychiatry 8:104–112. doi: 10.1177/002076406200800203

123. Sleath B, Svarstad B, Roter D (1998) Patient race and psychotropic prescribing during medical encounters. Patient Educ Couns 34:227–238.

124. Hunter LR, Schmidt NB (2010) Anxiety psychopathology in African American adults: Literature review and development of an empirically informed sociocultural model. Psychol Bull 136:211–235. doi: 10.1037/a0018133

125. Karasz A (2005) Cultural differences in conceptual models of depression. Soc Sci Med 60:1625–1635. doi: 10.1016/j.socscimed.2004.08.011

126. National Institute on Alcohol Abuse and Alcoholism (1998) Drinking in the United States: main findings from the 1992 National Longitudinal Alcohol Epidemiologic Survey (NLAES). National Institute on Alcohol Abuse and Alcoholism, Washington D.C.

127. Chatterji S, Saunders JB, Vrasti R, et al. (1997) Reliability of the alcohol and drug modules of the Alcohol Use Disorder and Associated Disabilities Interview Schedule -- Alcohol/Drug-Revised (AUDADIS-ADR): an international comparison. Drug Alcohol Depend 47:171–185.

128. Grant BF, Dawson DA, Stinson FS, et al. (2003) The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. Drug Alcohol Depend 71:7–16.

129. Grant BF, Stinson FS, Dawson DA, et al. (2004) Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Arch Gen Psychiatry 61:807–816.

130. Canino G, Bravo M, Ramirez R, et al. (1999) The Spanish Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS): reliability and concordance with clinical diagnoses in a Hispanic population. J Stud Alcohol 60:790–799.

131. Grant BF, Hasin DS, Stinson FS, et al. (2005) Co-occurrence of 12-month mood and anxiety disorders and personality disorders in the US: results from the national epidemiologic survey on alcohol and related conditions. J Psychiatr Res 39:1–9. doi: 10.1016/j.jpsychires.2004.05.004

132. Grant BF, Stinson FS, Hasin DS, et al. (2004) Immigration and lifetime prevalene of DSM-IV psychiatric disorders among Mexican Americans and non-Hispanic whites in the United States: results fromt he National Epidemiologic Survey on Alcohol and Related Conditions. JAMA Psychiatry 61:1226–1233.

133. Davies HTO, Crombie IK, Tavakoli M (1998) When can odds ratios mislead? Br Med J 316:989–991.

134. Ayanian JZ, Weissman JS, Chasan-Taber S, Epstein AM (1999) Quality of care by race and gender for congestiv heart failure and pneumonia. Med Care 37:1260–1269.

135. Ayanian JZ, Cleary PD, Weissman JS, Epstein AM (1999) The effect of patients' preferences on racial differences in access to renal transplantation. N Engl J Med 341:1661–1669.

136. Bach PB, Cramer LD, Warren JL, Begg CB (1999) Racial differences in the treatment of early-stage lung cancer. N Engl J Med 341:1198–1205.

137. Mayberry RM, Mili F, Ofili E (2000) Racial and ethnic differences in access to medical care. Med Care Res Rev 57:108–145.

138. Michaud CM, McKenna MT, Begg S, et al. (2006) The burden of disease and injury in the United States 1996. Popul Health Metr 4:11.

139. Kung HC, Hoyert DL, Xu J, Murphy SL (2008) Deaths: Final data for 2005. Natl Vital Stat Rep 56:1–120.

140. Mays VM, Cochran SD, Barnes NW (2007) Race, race-based discrimination, and health outcomes among African Americans. Annu Rev Psychol 58:201–225.

141. Williams DR, Mohammed SA (2008) Discrimination and racial disparities in health: evidence and needed research. J Behav Med 32:20–47. doi: 10.1007/s10865-008-9185-0

142. Brown C, Schulberg HC, Madonia MJ (1996) Clinical presentations of major depression by African Americans and whites in primary medical care practice. J Affect Disord 41:181–191.

143. Iwata N, Buka S (2002) Race/ethnicity and depressive symptoms: a cross-cultural/ethnic comparison among university students in East Asia, North and South America. Soc Sci Med 55:2243–2252.

144. Alegria M, Jackson JS, Kessler RC, Takeuchi D (2008) Collaborative Psychiatric Epidemiology Surveys (CPES), 2001-2003 [United States].

145. Kessler RC, Berglund P, Demler O, et al. (2003) The epidemiology of major depressive disorder. JAMA 289:3095–3105.

146. Angst J, Cui L, Swendsen J, et al. (2010) Major depressive disorder with subthreshold bipolarity in the National Comorbidity Survey Replication. Am J Psychiatry 167:1194–1201.

147. Furukawa TA, Kessler RC, Slade T, Andrews G (2003) The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. Psychol Med 33:357–362. doi: 10.1017/S0033291702006700

148. Andersen LS, Grimsrud A, Myer L, et al. (2011) The psychometric properties of the K10 and K6 scales in screening for mood and anxiety disorders in the South African Stress and Health study: Psychometric properties of K10 and K6 in South Africa. Int J Methods Psychiatr Res 20:215–223. doi: 10.1002/mpr.351

149. Barnes DM (manuscript in preparation) Testing a methodolgoical explanation for the Black-White depression paradox.

150. Coon HM, Kemmelmeier M (2001) Cultural orientations in the United States (re) examining differences among ethnic groups. J Cross-Cult Psychol 32:348–364.

151. Komarraju M, Cokley KO (2008) Horizontal and vertical dimensions of individualism-collectivism: A comparison of African Americans and European Americans. Cultur Divers Ethnic Minor Psychol 14:336–343. doi: 10.1037/1099-9809.14.4.336

152. Oyserman D, Coon HM, Kemmelmeier M (2002) Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. Psychol Bull 128:3–72. doi: 10.1037//0033-2909.128.1.3

153. Rosenfield S, Smith D (2009) Gender and mental health: do men and women have different amounts or types of problems? Handb. Study Ment. Health Soc. Contexts Theor. Syst., 2nd ed. Cambridge University Press, Cambridge, pp 256–267

154. King M, Semlyen J, Tai S, et al. (2008) A systematic review of mental disorder, suicide, and deliberate self harm in lesbian, gay and bisexual people. BMC Psychiatry 8:70. doi: 10.1186/1471-244X-8-70

155. Meyer IH (2013) Prejudice, Social Stress, and Mental Health in Lesbian, Gay, and Bisexual Populations. Psychol Sex Orientat Gend Divers 1:3–26.

156. Aneshensel CS (2009) Toward explaining mental health disparities. J Health Soc Behav 50:377–394.

157. Cochran SD, Sullivan JG, Mays VM (2003) Prevalence of mental disorders, psychological distress, and mental services use among lesbian, gay, and bisexual adults in the United States. J Consult Clin Psychol 71:53.

158. Mirowsky J, Ross CE (1986) Social patterns of distress. Annu Rev Sociol 12:23–45.

159. Helzer JE, Robins LN, McEvoy LT, et al. (1985) A comparison of clinical and diagnostic interview schedule diagnoses. Arch Gen Psychiatry 42:657–666.

Appendix A: Methodological notes on Chapter 2

<u>Literature Review</u>

My systematic literature review was limited to PubMed and PsycINFO databases, and did not include a review of grey literature, such as non peer-reviewed reports from government agencies. However, I supplemented the systematic review with articles and grey literature I had already collected. Between these sources, I was confident I would collect all eligible articles on major depression, given the limited number of eligible studies. I was also familiar enough with the distress findings to know that examples of Blacks having lower distress levels than Whites were rare and that my systematic review and existing store of articles would sufficiently capture the Black-White distress pattern.

<u>Exclusion of National Survey of American Life (NSAL) from depression findings</u>

Findings from the NSAL study, which was conducted concurrently with the National Comorbidity Survey Replication (NCS-R), were excluded in this systematic review because, despite meeting most eligibility criteria, the study only sampled Whites living in census tracts with an African-American population of at least 10 percent, which represents only 14 percent of Whites in the US [67]. Despite the fact that the data were then weighted to the US population by several demographic characteristics, they could not be weighted to account for this particular sampling feature. Nevertheless, in the NSAL, African Americans had a lower prevalence than Whites of lifetime, past year, and past 30-day major depression, and Caribbean Americans had a lower prevalence than Whites of lifetime and past 30-day depression [67]. These findings are consistent with the majority of findings reported in this paper. I am not aware of findings from this study comparing Blacks and Whites on distress.

<u>Estimating prevalence ratios and 95 percent confidence intervals for major depression in Table 2 and</u>

<u>high distress in Table 3a and 3b</u>

I estimated the prevalence ratios and confidence intervals in *openepi.com* where I entered each group's data in a 2x2 table. Figures for populating the 2x2 cells were derived from applying the Black and White prevalence figures reported in each study to the unweighted sample sizes. The *openepi.com* output from these data are risk ratios and their 95 percent confidence intervals (Taylor series); I call the former "prevalence ratios" in my review.

<u>Estimating t-tests to compare mean distress levels in Table 4</u>

I used *openepi.com* to conduct *t*-tests of Black-White differences in distress means whenever the sample sizes and either standard deviations or standard errors were provided in the papers.

Appendix B: Methodological notes on Chapter 3

Operationalizing somatization

I considered four options for operationalizing somatization: as a difference score between somatic and psychological symptoms endorsed, as a ratio of somatic to psychological symptoms endorsed, as a proportion of somatic symptoms to total symptoms endorsed, and as somatic symptoms and psychological symptoms endorsed as two separate continuous variables. In view of the strengths and weaknesses of each, on balance the first of these offered the best solution.

Somatization as a *ratio* of somatic to psychological items endorsed, or the reverse, entails a zero in the denominator in many cases, which is nonsensical and mathematically unworkable. Somatic items as a *proportion* of total items endorsed is conceptually problematic because a hypothetical person *A* endorsing 1 somatic symptom and no psychological symptoms would have a proportion score of 100%, whereas a hypothetical person *B* endorsing four somatic and one psychological symptom would have a proportion score of 80%. Person *B* has relatively high distress and we know that they express it relatively somatically. Person *A* has comparatively low distress and we do not know, based on this, how they would express distress at higher levels, but there is a chance that they would express it less somatically than person *B*. Therefore, it seems to lack face validity to assign a higher somatization score to person *A* than to person *B*. Somatic and psychological item endorsements as separate *continuous measures* is a problem because both measures would be strongly collinear with the measure of underlying depression (a pure symptom count across the seven symptoms) and all three measures would be included in the full model testing the second hypothesis. Alternatively, underlying depression could be omitted from the full model since separate measures of somatic and psychological symptom

endorsements would capture the same information. The weakness of this alternative is that a one-unit increment in the somatic score would have the same impact on the probability of endorsing a screening symptom regardless of the psychological measure score, which is a sub-optimal rendering of depression somatization, a point elaborated on below.

The difference measure does not have these weaknesses. As a non-ratio measure, it avoids the problem of having a zero in the denominator. As a non-proportion measure, it circumvents the specific problem of low face-validity highlighted in the previous paragraph.   It also avoids the problem of obvious collinearity with underlying depression introduced by creating separate measures for psychological and somatic symptom endorsements. The primary strength of the difference score is that the effect of a one-unit increase in somatic endorsements on the probability of endorsing a screening symptom *depends on* the level of psychological endorsements. By contrast, if somatization is modeled using two different variables, one for somatic symptom endorsement and one for psychological symptom endorsements, the effect of a one-unit increase in somatic endorsements on the probability of endorsing a screening symptom is *independent* of the number of psychological symptom endorsements. The advantage of *dependence* on psychological symptom endorsements is that depression somatization is conceptualized as the *relative* distribution of somatic to psychological symptom endorsements. Thus an increase in somatic symptom endorsements in an individual whose psychological symptom endorsements are low is more suggestive of a somatic expression of depression than when their psychological symptom endorsements are high, and this difference is hypothesized to bear on the probability of endorsing a psychological screening symptom.   The difference score captures this presumed difference, whereas separate variables for somatic and psychological symptom endorsements do not.

On the other hand, the relationship between the difference score measure and underlying depression is potentially over-determined because a person's underlying depression score and their psychological (or somatic) symptom endorsements determines their number of somatic (or psychological) symptom endorsements. Such over-determination could create collinearity between the underlying depression and difference score measures. To circumvent this problem, an internal committee member suggested the alternative described above of testing the second hypothesis by: 1) putting underlying depression in the first model, but not in the second model, and 2) accounting for both somatization and underlying depression in the second model by adding somatic symptom endorsements and psychological symptom endorsements as separate measures. This did not change the results, however. That is, the Black-White odds ratio for endorsing a screening symptom changed in the same direction and to the same small degree as in the original test of the second hypothesis modeling somatization using the difference score.

Also, Dr. Schwartz and I consulted with Dr. Melanie Wall in Columbia's Biostatistics department to discuss possible issues with having underlying depression and the somatization difference score in the same model. Dr. Wall said this would not be a problem because this was similar to conducting a principal components regression analysis.

<u>Age- and sex-standardization</u>

Black and Whites estimates in all analyses adjusted for the different age and sex distributions in the Black and White samples by standardizing to the White age-by-sex distributions in the NLAES sample. The table below shows the age-by-sex distributions in the Black (prior to standardization) and White NLAES samples.

Age	Sex	Black unweighted _n_	White unweighted _n_	Weighted Black %	Weighted White %
18-24	M	329	1,606	8.05	6.33
	F	528	1,755	9.22	6.09
25-44	M	884	5,978	21.95	21.41
	F	1,802	7,626	26.46	21.50
45-64	M	546	3,527	10.15	12.89
	F	855	4,398	12.52	13.70
65+	M	341	2,489	4.56	7.50
	F	670	4,559	7.11	10.58
Total		5,955	31,938	100.00	100.00

<u>Back-up to revised past-year depression prevalences applying the illness/bereavement exclusion to all</u>

In the past-year depression module, only those endorsing a screen symptom and three or more additional symptoms were asked the physical illness/bereavement exclusion question. In this sub-group, 58.24% of Blacks and 40.42% of Whites endorsed this exclusion. I applied these percents to those from the past-year module who received a depression diagnosis using the alternative algorithm, but who were not asked the exclusion question because they had not endorsed a screening question.

The weighted _n_'s of this latter group were:

Black: 193,642 x .5824 = 112,312

White: 876,385 x .4042 = 350,554

These resulting figures (112,312 and 350,554) were then subtracted from the weighted _n_'s who received a diagnosis using the alternative algorithm, to determine what the weighted _n_ would be had the physical illness/bereavement exclusion been applied to all. This assumes, of course, that the

endorsement patterns by race would be the same between those asked the exclusion question and those not asked it. The revised absolute prevalence figures for the alternative algorithm were:

Black: 719,941 – 112,312 = 607,629

White: 5,959,944 – 350,554 = 5,609,390

This translates into the following prevalence percentages:

Black: 607,629/20,642,209 = 2.94%

White: 5,609,390/140,115,224 = 4.00%

In turn, these percentages translate into a revised OR of 0.73, which is nearly the same as OR using the conventional algorithm.

Appendix C: Methodological notes on Chapter 4

Age- and sex-standardization

Black and Whites estimates in all analyses adjusted for the different age and sex distributions in the Black and White samples by standardizing to the White age-by-sex distributions in the NCS-R sample. The table below shows the age-by-sex distributions in the Black (prior to standardization) and White NCS-R samples.

Age	Sex	Black unweighted *n*	White unweighted *n*	Weighted Black %	Weighted White %
18-24	M	42	243	7.75	6.81
	F	78	264	12.27	5.92
25-44	M	118	738	18.34	16.87
	F	201	930	21.81	17.61
45-64	M	76	592	11.84	15.51
	F	141	801	16.28	17.83
65+	M	21	229	4.33	8.34
	F	40	383	7.38	11.11
Total		717	4,180	100.00	100.00