



Published in final edited form as:

Psychol Assess. 2011 September ; 23(3): 752–761. doi:10.1037/a0023288.

The utility of the Kessler Screening Scale for Psychological Distress (K6) in Two American Indian communities

Christina M. Mitchell and Janette Beals

Centers for American Indian and Alaska Native Health, Colorado School of Public Health, University of Colorado Denver.

Abstract

The Kessler Screening Scale for Psychological Distress (K6) has been used widely as a screener for mental health problems and as a measure of severity of impact of mental health problems. However, the applicability and utility of this measure for assessments within American Indian communities has not been explored. Data were drawn from a large-scale epidemiological study conducted in cooperation with two American Indian populations. Participants ($n = 3,084$) were 15 – 54 years old, living on or near their home reservations; each completed an interview that included a version of the Composite International Diagnostic Interview (CIDI) and the K6. Measures of both physical- and mental-health-related quality of life (the SF-36) were used to examine the importance of the K6 over and above psychiatric diagnoses. The K6 was shown to be an appropriate screening and severity measure for mood disorders in these two samples. It also predicted health-related quality of life over and above that predicted by diagnoses alone. Inclusion of a measure such as the K6 as a complement to more traditional dichotomous diagnoses in both research and clinical practice is recommended.

Keywords

Kessler Screening Scale for Psychological Distress; K6; American Indian

The utility of the Kessler Screening Scale for Psychological Distress (K6) in Two American Indian communities

A major advance in psychological research and epidemiology over the last 30 years has been the development of lay-administered structured interviews, allowing the standardized estimation of the prevalence of psychological disorders among non-clinical populations (Cairney, Veldhuizen, Wade, Kurdyak, & Streinter, 2007). In large-scale community-based epidemiological surveys such as the Epidemiological Catchment Area studies (Regier et al., 1990) and the National Comorbidity Surveys (Kessler, Berglund, Demler, Jin, & Walters, 2005; Kessler et al., 1994), nearly half of the general population have been found to qualify for one or more psychological disorders in their lifetimes; almost one in five had at least one diagnosis in the year prior to these surveys (Kessler, 2002). Such high rates of disorder lead

Correspondence concerning this article should be addressed to Christina M. Mitchell, Centers for American Indian and Alaska Native Health, Mail Stop F800, Nighthorse Campbell Native Health Building, 13055 E. 17th Avenue, Room 335, Aurora, Colorado 80045. Christina.Mitchell@UCDenver.edu.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/pas

to at least two important public policy questions: How can we most efficiently detect those with psychological disorders and, once identified, how best can we ascertain those in severe distress for whom scarce treatment resources are needed?

Nowhere are these issues more pressing than in American Indian communities. Past research among American Indian communities has revealed high rates of mental disorders, especially alcohol abuse/dependence and posttraumatic stress disorder (e.g., Kinzie et al., 1992; Kunitz et al., 1999; Robin, Chester, Rasmussen, Jaranson, & Goldman, 1997). Substantial tribal differences in the prevalence of certain problems, though—notably alcohol and drug use (e.g., May, 1996)—underscore the importance of including samples with social and cultural diversity. Recent studies have shown the burden of psychological disorders to be at least equivalent to, if not greater than, that of other Americans (Beals, Manson, et al., 2005; Beals, Novins, et al., 2005; Whitbeck, Hoyt, Johnson, & Chen, 2006). Yet at the same time, the treatment resources, especially within services funded by Indian Health Service (IHS), are severely curtailed (Roubideaux, 2005). In these populations, then, the identification of efficient methods of screening for those with likely disorders and determination of severity is especially important.

Screening

Structured diagnostic interviews such as the Composite International Diagnostic Interview (CIDI, Robins, Wing, Wittchen, & Helzer, 1988) assess multiple conditions using the often-complicated DSM-based nosologies. As a result, they can be time-consuming and burdensome to respondents (Cairney, et al., 2007). A two-stage sampling process, which screens for potential cases before a lengthy diagnostic interview is administered, may greatly reduce the number of structured interviews required. Screening measures have additional uses, ranging from serving as assessments of possible serious mental illness in population-based and clinical studies to the identification in busy primary care and clinical settings of those for whom a full diagnostic assessment is warranted (Kessler, 2002; Kessler et al., 2003; Kessler, et al., 2005). In American Indian communities, where services are scarce (Manson, 2001; Nelson, McCoy, Stetter, & Vanderwagen, 1992), an effective screening tool can be especially important for identifying those most in need of services. To minimize costs, such a screening instrument should be short and easy to administer. Also, an effective screener should not simply identify positive cases with a high degree of accuracy; to ensure that the first stage captures most of the probable cases, it should also eliminate a large proportion of negative cases. These aims are met by high values of sensitivity and specificity, respectively (Cairney, et al., 2007).

Severity

In addition to the concern of efficiently screening for disorder, identifying the degree of severity of a disorder's impact is important. Severity has been found to be a significant predictor of a broad range of clinical phenomena, including comorbidity, role impairment, and the course and chronicity of disorder (Furukawa, Kessler, Slade, & Andrews, 2003; Watson, 2005). Clinically, people with the same diagnosis can vary enormously in the severity of their illness (Clark, Watson, & Reynolds, 1995; Kessler, 2002). Indeed, many of those diagnosed with disorder in the community had considerably less severe disorders than did those who had sought treatment (Heath et al., 1994; Kessler, 2002). Thus, estimates of service need founded on population-based estimates of disorder can be strengthened with severity assessments.

The Kessler Screening Scale for Psychological Distress (K6)

One of the most widely used measures, for either screening or severity, is the six-item K6 screening scale for psychological distress, developed by Kessler and colleagues (Kessler et

al., 2002), along with the earlier 10-item K10, as a measure of nonspecific psychological distress. More specifically, using a 30-day reference period, respondents rate how often they felt nervous, hopeless, restless or fidgety, so sad that nothing could cheer them up, that everything was an effort, and worthless (Fleishman & Zuvekas, 2007; Furukawa, et al., 2003). In assessing the utility of the K6, two types of studies are common. First, the ability of the K6 to predict DSM-based diagnoses has been explored in a wide range of samples (Baggaley et al., 2007; Furukawa, et al., 2003; Veldhuizen, Cairney, Kurdyak, & Streiner, 2007). Second, investigators have examined the degree to which the K6 is related to known correlates of severity (Fleishman & Zuvekas, 2007; Kessler, et al., 2003; Swartz & Lurigio, 2006). Health-related quality of life is used as a common indicator of severity, such as that assessed by the Medical Outcome Study's Short Form (SF) -12 and SF-36 (Andrews & Slade, 2001; Fleishman & Zuvekas, 2007; Gill, Butterworth, Rodgers, & Mackinnon, 2007). The outcomes of these efforts have indicated that the K6 is both an effective screening measure and an indicator of distress severity among the populations that have participated to date. Indeed, Kessler and others have advocated for the inclusion of this measure as a standardized measure in those contexts where fuller assessments of disorder and severity are not possible (Kessler, et al., 2002).

Study Goals

Using data from a large epidemiological study of two tribal groups of American Indians living on or near their reservations, we examined two research questions. First, is the K6 psychometrically appropriate for use in this population? We used receiver operating characteristics analyses to examine its utility to predict DSM-IV mood disorders as defined by a lay-administered structured interview. Second, how might the K6 add information for researchers and clinicians about the severity of a diagnosis that moves beyond the presence/absence of emotional, substance use, and physical disorders? In other words, what is the incremental validity of the K6, over and above diagnoses?

Methods

Sample

The primary objective of the American Indian Service Utilization, Psychiatric Epidemiology, and Risk and Protective Factors Project (AI-SUPERPFP) was to estimate the prevalence of psychiatric disorders and associated service utilization in two American Indian reservation populations. The populations of inference were defined as 15- to 54-year-old enrolled members of two closely related Northern Plains (NP) tribes and a Southwestern tribe (SW) living on or within 20 miles of their reservations at the time of sampling (1997). In order to protect the confidentiality of the tribal communities involved in this research (Norton & Manson, 1996), we refer to the tribes using these general descriptors (NP and SW) rather than specific tribal names.

The two participating tribal groups represent both the diversity and common experiences of this population. They belong to different linguistic families, have different histories of migration, subscribe to different principles for reckoning kinship and residence, and have historically pursued different forms of subsistence. Yet both tribes have many experiences in common with other American Indian groups: similar histories of colonization, externally imposed forms of governance, forced dietary changes, and mandatory boarding school education. Unemployment was widespread, ranging from 40% to 90% of the eligible labor force. Both tribes represent considerable variability in acculturation, education, and income.

Tribal rolls formed the sampling universe; these records list all individuals meeting minimal requirements for recognition as tribal members. A critical point for AI-SUPERPFP was the

fact that tribal enrollment coincided with eligibility for IHS services—the major health care provider in the rural communities involved and a major focus of our services research.

Stratified random sampling procedures were used; tribe, age (15–24, 25–34, 35–44, and 45–54 years), and gender served as strata. Records were selected randomly for inclusion into replicates, which were then released as needed to reach the goal of approximately 1,500 interviews per tribe. An elaborate location procedure was developed, including searches of public records and queries of family members and knowledgeable community “key informants” by local tribal members who were employed in the Field Office by the university; supervisors rather than interviewers made the final location determination. Altogether 46.6% and 39.2% of those listed in the SW and NP tribal rolls were found to be living on or near their reservations. Once determined eligible, Field Office staff returned to describe and request participation in the project. Of those located and found eligible, 76.8% in the NP (N=1,638) and 73.7% in the SW (N=1,446) agreed to participate. Sample weights accounted for differential selection probabilities across all strata and for non-response biases. The AI-SUPERPPF methods are described in greater detail elsewhere (Beals, Manson, Mitchell, Spicer, & AI-SUPERPPF Team, 2003); our Web site (<http://www.ucdenver.edu/academics/colleges/PublicHealth/research/centers/CAIANH/NCAIANMHR/ResearchProjects/Pages/AI-SUPERPPF.aspx>) provides additional detail, including copies of the interview and the training manual.

Data Collection

Tribal and university Institutional Review Board approvals were obtained prior to data collection. All adult participants provided informed consent; parental/guardian consent was obtained before requesting adolescent assent. Interviewers were recruited through advertisements placed in local and Native newspapers. Since few of the Field Office staff had more than a high school education, intensive training and quality control procedures provided the skills necessary to yield reliable and valid findings. Details about the training and greater specificity about data collection may be found in the training manual available on our Web site (<http://www.ucdenver.edu/academics/colleges/PublicHealth/research/centers/CAIANH/NCAIANMHR/ResearchProjects/Documents/manual.pdf>); however, several modifications to usual practice are important to mention. The training was longer than typical efforts and included two weeks of didactic instruction followed by at least another week of practice sessions. Periodic refresher courses and quality control monitoring continued throughout, conducted independently by both local field supervisors and staff from our Denver office.

Measures

Measures were reviewed by community members in focus groups prior to the beginning of data collection for issues of cultural appropriateness and cultural relevance (Beals, et al., 2003). A focus group format was used with 4 groups: male and female tribal members, service providers, and elder. Each section of the interview was reviewed by at least one of the first three groups; elders were asked to comment more broadly about what issues were most important to assess. Of the measures used here, the CIDI was the only one adapted, as noted below.

K6—As noted earlier, the K6 (Kessler, et al., 2002) contains 6 questions that ask about the following feelings during the past month: sad, nervous, restless or fidgety, hopeless, everything is an effort, worthless. No concerns about the cultural validity of this measure were raised. Cronbach’s alpha for the full sample for this scale was .83; a one-factor confirmatory factor analysis (comparative fit index = .95) demonstrated a satisfactory fit of

all 6 items onto one underlying factor. Responses range from 0 (none of the time) to 4 (all of the time). We summed across these 6 items for a scale score ranging from 0 – 24.

Diagnoses—DSM-IV disorders were measured with the University of Michigan version of the CIDI (Kessler, et al., 1994). The CIDI was adapted for use in American Indian communities in the context of a previous project (Beals, et al., 2003) with judicious use of some rewording; the most common strategy was to add additional questions where concerns were raised that a question may be misunderstood. For example, focus group members suggested that, in the Major Depressive Episode section, we include the question about irritability, which typically is asked of adolescents but not adults, since they felt it could be common expression for depression for adults in these communities as well. The AI-SUPERPFP CIDI yielded diagnoses of the following disorders: major depressive episode (MDE), dysthymic disorder, generalized anxiety disorder (GAD), panic disorder, posttraumatic stress disorder (PTSD), alcohol use disorders (abuse and dependence combined), and drug use disorders (abuse or dependence for any of marijuana, cocaine, inhalant, hallucinogens, heroin, sedative, tranquilizer, stimulant, analgesic use disorders).

We used the diagnostic variables in two ways. First, to examine the psychometric properties of the K6 as a screener, we created three composite variables of past-year diagnosis: depression/dysthymia, any anxiety disorder (GAD, panic, and/or PTSD), and any mood disorder (any depressive or anxiety disorder) (Baillie, 2005; Cairney, et al., 2007; Furukawa, et al., 2003). In these analyses, we focused only on past-year disorders as more appropriate than lifetime disorders, given the K6's reference period of the past 30 days. (Past-month disorders were also available. However, despite the fact that AI-SUPERPFP was the largest study of its kind in American Indian communities, those prevalence rates were too low for this analysis.)

Second, to examine the K6's incremental validity and potential clinical utility, we used both mood and alcohol use disorders. (It should be noted that few cases existed of drug use disorder independent of alcohol use disorder; thus, only the latter was used.) For both types of disorder, two variables were created: disorder in the past year and previous disorder during lifetime but not the past year. It was anticipated that disorders in the past year would be most clearly linked to health-related quality of life at time of interview but, further, that a lifetime history of disorder may also have an association.

Outcomes for the incremental validity analyses: Short Form (SF)-36—As one of the most commonly used outcome measures used in clinical settings (McHorney, Ware, Lu, & Raczek, 1993; McHorney, Ware, & Raczek, 1993; Ware & Sherbourne, 1992), the SF-36 provided an indicator of current functioning. Based on focus group feedback, minor modifications were made to the SF-36; mostly in terms of alternative activities (e.g., substituting “horse-back riding” for “golf” as a moderate level activity); a practice encouraged by the developers of the SF-36. The literature has shown that those with more severe psychological disorders or emotional problems are more likely than their healthier counterparts to report lower health-related quality of life in both the mental and physical health realms (Andrews & Slade, 2001; Gill, et al., 2007). Thus, the outcomes used in the incremental validity analyses were the Physical Component and Mental Component Summaries (PCS and MCS, respectively) of the SF-36, reflecting the common finding that the eight subscales assessed in the SF-36 are adequately explained by this underlying two-factor structure (Ware, 1994).

Control variables—We used three *demographic* variables as controls, as others have (e.g., Baillie, 2005; Swartz & Lurigio, 2006)—age, gender, and ethnicity (here, tribe). *Age* was a continuous variable of age in years. *Gender* was operationalized as female gender (0 =

male, 1 = female); *tribe* was coded as NP (0 = SW, 1 = NP). We also included a multiplicative interaction term of gender by tribe.

In order to adequately assess the ability of the K6 to predict the PCS measure derived from the SF-36, it was important to include a measure of *physical health*. AI-SUPERPPF included a list of 31 possible chronic or debilitating physical conditions reported by the participant to have been diagnosed by a physician. Parallel to the psychological diagnoses, we created two variables: 1) history of any physical diagnosis during lifetime but not past year and 2) any physical health diagnosis in the past year.

Analytic Strategies

Psychometric analyses—A common approach to determining the utility of a scalar screening measure is to compare it to a “gold standard”—here, contrasting the K6 with a DSM diagnosis—using receiver operating characteristics (ROC) analyses. Two aspects are key to this analysis: sensitivity (true positives, or those determined by the gold standard to have a diagnosis who are also screened as having this disorder using the scalar measure) and specificity (true negatives, or those without a disorder who are also found not to have this disorder using the scalar measure). Traditionally, the ROC curve plots true positives (sensitivity) against false positives (one minus specificity) at each score of the continuous measure.

With any screener, the distributions of scores for those with and without a diagnosis will overlap to some extent (Hanley & McNeil, 1982); the challenge is to determine whether a screener can differentiate those with and without a diagnosis sufficiently accurately. The area under the curve (AUC) is the primary measure of the accuracy of such measures, representing the probability that a randomly selected person *with* a diagnosis will have a higher score on the scalar measure than will a randomly selected person *without* a diagnosis (Cairney, et al., 2007). Thus, the AUC is a quantitative assessment of the congruence between the scalar measure and the standard. If the sensitivity and specificity values track across the line of no information—a straight line stretching from the lower left corner to the upper right corner—the measure cannot discriminate those with a diagnosis from those without. The more the curve arches toward the upper left corner, the better the screener (Tsuang, Tohen, & Zahner, 1995). The AUC ranges from .5 for a measure with no diagnostic power to 1 for a perfect measure (Johnson, 2004). In general, AUCs between 0.5 and 0.7 are considered to reflect low accuracy; those between 0.7 and 0.9, moderate accuracy; and those between 0.9 and 1.0, high accuracy (Cairney, et al., 2007).

Incremental validity—To examine the second question of the incremental validity and, therefore, possible clinical utility of the K6, we used Mplus, Version 5 (Muthén & Muthén, 1998–2007) to conduct all regressions. The dependent variables were the PCS and MCS scales of the SF-36, modeled simultaneously in a multivariate multiple regression to take into account the correlation between the two dependent variables. Mplus uses full-information maximum likelihood estimation to incorporate all available information; as a result, no special handling of missing data was required. All models included age, gender, tribe, and a gender-by-tribe interaction as control variables.

Results

Sample Descriptions

As detailed elsewhere (Beals, Manson, et al., 2005), the AI-SUPERPPF samples adequately represented those living on or near the reservations in question at the time of sample development (1997). Demographically, substantially more women than men were

interviewed in the SW, likely reflecting differential migration patterns where men were more likely to pursue employment in off-reservation urban areas (see Table 1). SW participants were more likely than their NP counterparts to be married and less likely to be separated, divorced, or widowed.

Few sample differences were detected in the prevalence of mood disorders, with the exception of PTSD: Both samples of women were more likely to qualify for this diagnosis than were men. For alcohol use disorders, the SW women were less likely to qualify for diagnosis than were all other samples; NP women were more likely to qualify for alcohol use disorders than were SW women but they were less likely to have alcohol use disorders than were either male sample.

Psychometric Analyses

AUC—We conducted 5 ROC analyses (total sample, by gender, and by tribe) for the three diagnostic “gold standards” for past-year disorder: depression or dysthymia, any anxiety disorder, and any mood disorder. AUC results and 95% confidence intervals are presented in Table 2; the ROC curves are displayed in Figures 1 – 3. As noted earlier, the AUC assesses the congruence between the continuous measure and the standard. For the total sample, the AUCs ranged from .73 (any anxiety disorder) to .83 (depression/dysthymia). Looking at 95% confidence intervals, the AUC for depression/dysthymia in the total sample was significantly higher than for any anxiety disorder or any mood disorder. None of the AUC estimates *within* diagnosis, by gender or tribe, differed significantly.

Heterogeneity within diagnostic group—We examined the distributions of the K6 scores across disorder groupings. The most conservative examination compared those with and without any mood disorder, with mood disorders including a heterogeneous combination of disorders. The boxplot in Figure 4 displays the spread of K6 scores for the “no diagnosis” group (0) and the group with any mood disorder in the past year (1). The heavy line represents the median of the group; the box around the heavy line, those lying within one quartile of the median; the bottom and top horizontal lines, the range of those within three quartiles. This figure shows that the distributions of K6 scores for these two groups had some overlap: The group with no past-year diagnosis had a median K6 score of 2 and a range of 0 – 21, with 26 outliers falling above the 3rd quartile (not shown); the group with a past-year diagnosis had a median K6 score of 7, with a range of 0 – 24. However, given that the interquartile box of the group with a disorder was twice the length of that of the group with a disorder, this plot even more dramatically shows the heterogeneity of K6 scores within the group with a past-year mood disorder.

Incremental Validity

We examined three models of incremental validity of the K6. (See Table 3.) A first test is its ability to predict severity of the disorder’s impact over and above its most closely related diagnostic category—past-year mood disorder. In Model 1, past-year mood disorder was significantly related to not only the MCS in the expected direction but also the PCS: Those with a past-year mood disorder reported lower quality of life both physically and mentally. Moreover, the K6 was not only a significant predictor, but its standardized parameters were 2.5 times larger than the standardized parameters of past-year mood disorder.

We next added previous lifetime (but not past-year) mood disorder as a predictor (Model 2). Although lifetime mood disorder was a significant predictor of both PCS and MCS in the expected direction, its inclusion in this model produced little change in the parameters of either past-year mood disorder or the K6: The K6 standardized parameters still remained more than twice the size of those of past-year mood disorder.

Finally, we added alcohol use disorder (lifetime/not past-year and past-year) and physical diagnosis (lifetime/not past-year and past-year; Model 3). With the inclusion of these new variables, previous lifetime mood disorder remained a significant predictor only of MCS, although past-year mood disorder was significant for both PCS and MCS. Previous lifetime alcohol use disorder was not a significant predictor of either dependent variable; past-year alcohol use disorder was significantly related to MCS only. Lifetime physical diagnosis was related to neither dependent variable, although past-year physical diagnosis was a significant predictor of both PCS and MCS. Even with this final stringent test of the incremental validity of the K6, it remained a significant predictor of both PCS and MCS, with the standardized parameters remaining more than 2.5 times greater than those of past-year mood disorder.

Discussion

This study explored two questions. First, is the K6 an appropriate screening measure in these two tribal populations? Using three past-year diagnoses—depression/dysthymia, any anxiety disorder, and any mood disorder—the ROC analyses and the AUCs were not dramatically different from those found in other samples, falling in the moderate category. For instance, using the National Household Survey on Drug Abuse, Kessler et al. (Kessler, et al., 2003) found that the K6 had an AUC of .86 predicting any DSM-IV disorder except a substance use disorder. Among more than 36,000 Canadians, Cairney et al. (Cairney, et al., 2007) found that the K6 had an AUC of .86 predicting past-year depression. These findings suggest that the K6 can function as a general indicator of possible psychological disorder among American Indians of these tribes. Thus, it could serve as the first stage in a two-stage sampling process that screens for potential cases before a lengthy diagnostic interview is administered or as a screener in a primary care physician's or clinician's office.

In other samples, the K6 has been used to estimate prevalence of disorders by implementing a cut-off score of 13 or greater to indicate the presence of disorder. Among these American Indian samples, for instance, 94% of the cases of depression or dysthymia would be correctly classified using this cut-off score. As is true in any community-based study, though, this high percentage was largely driven by the vast number of non-cases correctly identified: Sensitivity (true positives) was 29% while sensitivity (true negatives) was 96%. If the cut-off score for the K6 were to be used in this way, it would be important to consider the context in which the decision criterion will be utilized. For instance, for research purposes, one may prefer a lower cut-off score in order to have a greater probability of true "caseness." Yet the heterogeneity of the severity of the impact of disorder on people's lives was clear. Thus, if one were recommending referral to extremely limited resources such as clinical services, one might want to use a higher cut-off score, to ensure that those who were more seriously affected received services.

The second question focused on the incremental validity and potential clinical utility of the K6 over and above other diagnostic measures. Even in the most conservative model (Model 3), where both previous lifetime and past-year diagnoses of mood disorders, alcohol use disorders, and physical diagnoses, the K6 consistently provided significant predictive information about physical and mental quality of life. Thus, the K6 appears to provide information about the severity and individuals' distress that is unique beyond mood, substance, and physical disorders and thus could be important for clinicians to be aware of.

Limitations

Recognizing the limitations of a study is critical. Here, we focused on only two of the larger tribal populations in the country and, of these, only members living on or near their reservations. In the final analysis, the AI-SUPERPFP samples, while well defined and

justified, were limited in cultural representation, age range, and residence. Thus, extrapolations to other groups should be made carefully. However, as a quick, simple, and straightforward assessment tool, the K6 appears to offer promise for understanding the impact of psychological disorder on people's everyday lives and for making treatment decisions accordingly.

Specific to these analyses, the cut-off scores used here (i.e., $.5 - .7$ = low accuracy, $.7 - .9$ = moderate accuracy, $> .9$ = high) are common in studies of the performance of psychiatric screeners. However, we acknowledge that the determination of the such cut-off scores is influenced by a number of factors, such as the adequacy of the "gold standard," the independence of this standard from the methods used, and sample characteristics (Swets). In addition, the K6 used the time reference of "past month," while we used past-year disorder to assess its psychometric capabilities. Past-month disorder would have been a closer approximation; as noted earlier. Yet while AI-SUPERPFP assessed past-month disorder, the prevalence rates were very low and could not be used here. Thus, the ROC analyses comparing the K6 with the DSM gold standard were definitely an *underestimate* of its comparability.

A possible confound could have arisen from the fact that the K6 and one of the 5 subscales of the MCS (Mental Health) have some overlap in questions: e.g., "How much of the time in the past 4 weeks have you felt so down in the dumps that nothing could cheer you up" (MCS) and "In the past 30 days, about how often did you feel so sad that nothing could cheer you up?" (K6). This overlap could have accounted for the stronger predictive power of the K6 with the MCS than with the PCS. To check this, we created a composite MCS variable that excluded the Mental Health subscale and reran the regressions. The two variable were correlated $.89$; as a result, the regressions using the shortened composite produced patterns identical to those using the full MCS score. This finding supported the fact that the predictive power of the K6 was not explained solely by the overlap in item content with the MCS. Because of this, we presented the analyses using the typical configuration of the MCS.

Finally, we remind readers that the SF-36 Component Summary measures (the PCS and MCS) were utilized here. These weighted summary measures, while convenient, must not be used uncritically. In particular, the weighting approach is based on normative U.S.-population-based samples and may not be appropriate for specific sub-populations or health groupings (Beals, Welty, Mitchell et al., 2006; Simon, Revicki, Grothaus, & Vonkorff, 1998; Taft, Karlsson, & Sullivan, 2001). Preliminary analyses determined that, for the samples here, the standard approach was adequate. Further, the psychometrically driven approach that yields orthogonal assessments of physical and mental health is troublesome. For the sake of comparability, the analyses presented here were modeled on those of others; we encourage further discussions of the appropriate uses of the SF-36 and related measures.

Conclusions

The advantages of utilizing a scalar measure, either for screening or for assessing severity, are clear. Such measures rely on answers representing a range of responses, as compared to the dichotomous assessment of the presence or absence disorder. In research contexts, scalar measures, by their very nature, tend to be more stable over time and display higher levels of reliability than do dichotomous measures—in large part because scalar scores are largely unaffected by relatively minor shifts in symptomatology (Watson, 2005). In addition, since so much variance is masked by a single categorical distinction, much larger samples are necessary to achieve the same statistical power in looking for important or explanatory relationships (Helzer et al., 2006; Kessler, 2002)

For clinical purposes, the utility and indeed necessity of retaining the dichotomous diagnostic approach is unquestioned (First & Westen, 2007). However, the debate about the clinical importance of such a measure has recently intensified. Currently, discussions are under way in preparation for the next version of the DSM—the DSM-V—and a number of points have been raised. For instance, even given the utility of a dichotomous diagnosis, clinicians also consider the severity of an illness in determining treatment recommendations (First & Westen, 2007). More broadly, severity of illness has significant implications for treatment, prognosis, and etiology (Klein, 2008). Indeed, DSM-IV has a severity specifier, but it can be difficult to use (First & Westen, 2007; Klein, 2008). Our findings add to those voices calling for both approaches. The inclusion of a scalar assessment such as the K6 can be an important complement to more traditional dichotomous decisions of presence or absence of disorder in more thoroughly understanding the impact of psychological disorders on daily life and in subsequently making more informed treatment recommendations and plans.

Acknowledgments

Data collection was supported by National Institute of Mental Health grants R01 MH48174 (Manson and Beals, PIs) and P01 MH42473 (Manson, PI); data analyses and writing, by R01 MH073965 (Beals, PI) and R01 MH075831 (Kaufman, PI).

AI-SUPERPPF would not have been possible without the significant contributions of many people. The following interviewers, computer/data management and administrative staff supplied energy and enthusiasm for an often difficult job: Anna E. Barón, Antonita Begay, Amelia T. Begay, Cathy A.E. Bell, Phyllis Brewer, Nelson Chee, Mary Cook, Helen J. Curley, Mary C. Davenport, Rhonda Wiegman Dick, Marvine D. Douville, Pearl Dull Knife, Geneva Emhoolah, Fay Flame, Roslyn Green, Billie K. Greene, Jack Herman, Tamara Holmes, Shelly Hubing, Cameron R. Joe, Louise F. Joe, Cheryl L. Martin, Jeff Miller, Robert H. Moran Jr., Natalie K. Murphy, Melissa Nixon, Ralph L. Roanhorse, Margo Schwab, Jennifer Settlemyre, Donna M. Shangreaux, Matilda J. Shorty, Selena S. S. Simmons, Wileen Smith, Tina Standing Soldier, Jennifer Truel, Lori Trullinger, Arnold Tsinajinnie, Jennifer M. Warren, Intriga Wounded Head, Theresa (Dawn) Wright, Jenny J. Yazzie, and Sheila A. Young. We would also like to acknowledge the contributions of the Methods Advisory Group: Margarita Alegria, Evelyn J. Bromet, Dedra Buchwald, Peter Guarnaccia, Steven G. Heeringa, Ronald Kessler, R. Jay Turner, and William A. Vega. Finally, we thank the tribal members who so generously answered all the questions asked of them.

References

- Andrews G, Slade T. Interpreting scores on the Kessler Psychological Distress Scale (K10). *Australian and New Zealand Journal of Public Health*. 2001; 25(6):494–497. [PubMed: 11824981]
- Baggaley RF, Ganaba R, Filippi V, Kere M, Marshall T, Sombie I, et al. Detecting depression after pregnancy: The validity of the K10 and K6 in Burkina Faso. *Tropical Medicine and International Health*. 2007; 12(10):1225–1229. [PubMed: 17956505]
- Baillie AJ. Predictive gender and education bias in Kessler's psychological distress scale (K10). *Social Psychiatry and Psychiatric Epidemiology*. 2005; 40:743–748. [PubMed: 16142511]
- Beals J, Manson SM, Mitchell CM, Spicer P. AI-SUPERPPF Team. Cultural specificity and comparison in psychiatric epidemiology: Walking the tightrope in American Indian research. *Culture, Medicine, and Psychiatry*. 2003; 27:259–289.
- Beals J, Manson SM, Whitesell NR, Spicer P, Novins DK, Mitchell CM. Prevalence of DSM-IV disorders and attendant help-seeking in two American Indian reservation populations. *Archives of General Psychiatry*. 2005; 62:99–108. [PubMed: 15630077]
- Beals J, Novins DK, Whitesell NR, Spicer P, Mitchell CM, Manson SM. Prevalence of mental disorders and utilization of mental health services in two American Indian reservation populations: Mental health disparities in a national context. *American Journal of Psychiatry*. 2005; 162:1713–1722. [PubMed: 16135632]
- Cairney J, Veldhuizen S, Wade TJ, Kurdyak P, Streinter DL. Evaluation of 2 measures of psychological distress as screeners for depression in the general population. *Canadian Journal of Psychiatry*. 2007; 52:111–120.

- Clark LA, Watson D, Reynolds S. Diagnosis and classification of psychopathology: Challenges to the current system and future directions. *Annual Review of Psychology*. 1995; 46:121–153.
- First MB, Westen D. Classification for clinical practice: How to make ICD and DSM better able to serve clinicians. *International Review of Psychiatry*. 2007; 19(5):473–481. [PubMed: 17896228]
- Fleishman JA, Zuvekas SH. Global self-rated mental health: Associations with other mental health measures and with role functioning. *Medical Care*. 2007; 45(7):602–609. [PubMed: 17571008]
- Furukawa TA, Kessler RC, Slade T, Andrews G. The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychological Medicine*. 2003; 33:357–362. [PubMed: 12622315]
- Gill SC, Butterworth P, Rodgers B, Mackinnon A. Validity of the mental health component scale of the 12-item Short-Form Health Survey (MCS-12) as measure of common mental disorders in the general population. *Psychiatry Research*. 2007; 152(1):63–71. [PubMed: 17395272]
- Hanley JA, McNeil BJ. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. *Diagnostic Radiology*. 1982; 143(1):29–36.
- Heath AC, Bucholz KI, Slutske WS, Madden PAF, Dinwiddie SH, Dunne MP, et al. The assessment of alcoholism in surveys of the general community: What are we measuring? Some insights from the Australian twin panel interview survey. *International Review of Psychiatry*. 1994; 6:295–307.
- Helzer JE, Bucholz KK, Beierut LJ, Regier DA, Schuckit MA, Guth SE. Should DSM-V include dimensional diagnostic criteria for alcohol use disorders? *Alcoholism: Clinical and Experimental Research*. 2006; 30(2):303–310.
- Johnson MP. Advantages to transforming the receiver operating characteristics (ROC) curve into likelihood ratio co-ordinates. *Statistics in Medicine*. 2004; 23:2257–2266. [PubMed: 15236429]
- Kessler RC. The categorical versus dimensional assessment controversy in the sociology of mental illness. *Journal of Health and Social Behavior*. 2002; 43(2):171–188. [PubMed: 12096698]
- Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*. 2002; 32(6):959–976. [PubMed: 12214795]
- Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, et al. Screening for serious mental illness in the general population. *Archives of General Psychiatry*. 2003; 60(2):184–189. [PubMed: 12578436]
- Kessler RC, Berglund P, Demler O, Jin R, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*. 2005; 62(6):593–602. [PubMed: 15939837]
- Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. *Archives of General Psychiatry*. 1994; 51:8–19. [PubMed: 8279933]
- Kinzie JD, Leung PK, Boehnlein J, Matsunaga D, Johnston R, Manson SM, et al. Psychiatric epidemiology of an Indian village: A 19-year replication study. *Journal of Nervous and Mental Disorder*. 1992; 180:33–39.
- Klein DN. Classification of depressive disorders in the DSM-V: Proposal for a two-dimension system. *Journal of Abnormal Psychology*. 2008; 117(3):552–560. [PubMed: 18729608]
- Kunitz SJ, Gabriel KR, Levy JE, Henderson E, Lampert K, JMcCloskey J, et al. Alcohol dependence and conduct disorder among Navajo Indians. *Journal of Studies on Alcohol*. 1999; 60:159–167. [PubMed: 10091952]
- Manson, SM. Behavioral health services for American Indians: Need, use, and barriers to effective care. In: Dixon, M.; Roubideaux, Y., editors. *Promises to keep: Public health policy for American Indians and Alaska Natives in the 21st century*. Washington, DC: American Public Health Association; 2001. p. 167-192.
- May, PA. Overview of alcohol abuse epidemiology for American Indian populations. In: Sandefur, GD.; Rindfuss, RR.; Cohen, B., editors. *Changing numbers, changing needs: American Indian demography and health*. Washington, D.C: National Academy Press; 1996. p. 235-261.

- McHorney CA, Ware JE Jr, Lu R, Raczek AE. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care*. 1993; 32(1):40–66. [PubMed: 8277801]
- McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care*. 1993; 31(3):247–263. [PubMed: 8450681]
- Muthén, B.; Muthén, L. *Mplus User's Guide*. 5th edition. Los Angeles: Muthén & Muthén; 1998–2007.
- Nelson SH, McCoy GF, Stetter M, Vanderwagen WC. An overview of mental health services for American Indians and Alaska Natives in the 1990s. *Hospital and Community Psychiatry*. 1992; 43:257–261.
- Norton IM, Manson SM. Research in American Indian and Alaska Native communities: Navigating the cultural universe of values and process. *Journal of Consulting and Clinical Psychology*. 1996; 64:856–860. [PubMed: 8916611]
- Regier DA, Farmer ME, Rae DS, Locke BZ, Keith SJ, Judd LL, et al. Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study. [see comments]. *Journal of the American Medical Association*. 1990; 264(19):2511–2518. [PubMed: 2232018]
- Robin RW, Chester B, Rasmussen JK, Jaranson JM, Goldman D. Prevalence and characteristics of trauma and posttraumatic stress disorder in a southwestern American Indian community. *American Journal of Psychiatry*. 1997; 154(11):1582–1588. [PubMed: 9356568]
- Robins LN, Wing J, Wittchen H, Helzer JE. The Composite Diagnostic Interview: An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry*. 1988; 45:1069–1077. [PubMed: 2848472]
- Roubideaux Y. Beyond Red Lake -- The Persistent Crisis in American Indian Health Care. *New England Journal of Medicine*. 2005; 353(18):1881–1883. [PubMed: 16267317]
- Swartz JA, Lurigio AJ. Screening for serious mental illness in populations with co-occurring substance use disorders: Performance of the K6 scale. *Journal of Substance Abuse Treatment*. 2006; 31:287–296. [PubMed: 16996391]
- Swets JA. Measuring the accuracy of diagnostic systems. *Science*. 1988; 240:1285–1293. [PubMed: 3287615]
- Tsuang, MT.; Tohen, M.; Zahner, GEP., editors. *Textbook in psychiatric epidemiology*. New York: Wiley-Liss; 1995.
- Veldhuizen S, Cairney J, Kurdyak P, Streiner DI. The sensitivity of the K6 as a screen for any disorder in community mental health surveys: A cautionary note. *Canadian Journal of Psychiatry*. 2007; 52:256–259.
- Ware, JE, Jr. *SF-36 physical and mental health summary scales: A user's manual*. Boston, MA: The Health Institute; 1994.
- Ware JE Jr, Sherbourne CD. The MOS 36-item Short Form Health Survey (SF-36): I. Conceptual framework and item selection. *Medical Care*. 1992; 30:473. [PubMed: 1593914]
- Watson D. Rethinking the mood and anxiety disorders: A quantitative hierarchical model for DSM-V. *Journal of Abnormal Psychology*. 2005; 114(2):522–536. [PubMed: 16351375]
- Whitbeck LB, Hoyt D, Johnson K, Chen X. Mental disorders among parents/caretakers of American Indian early adolescents in the Northern Midwest. *Social Psychiatry and Psychiatric Epidemiology*. 2006; 41(8):632–640. [PubMed: 16779502]

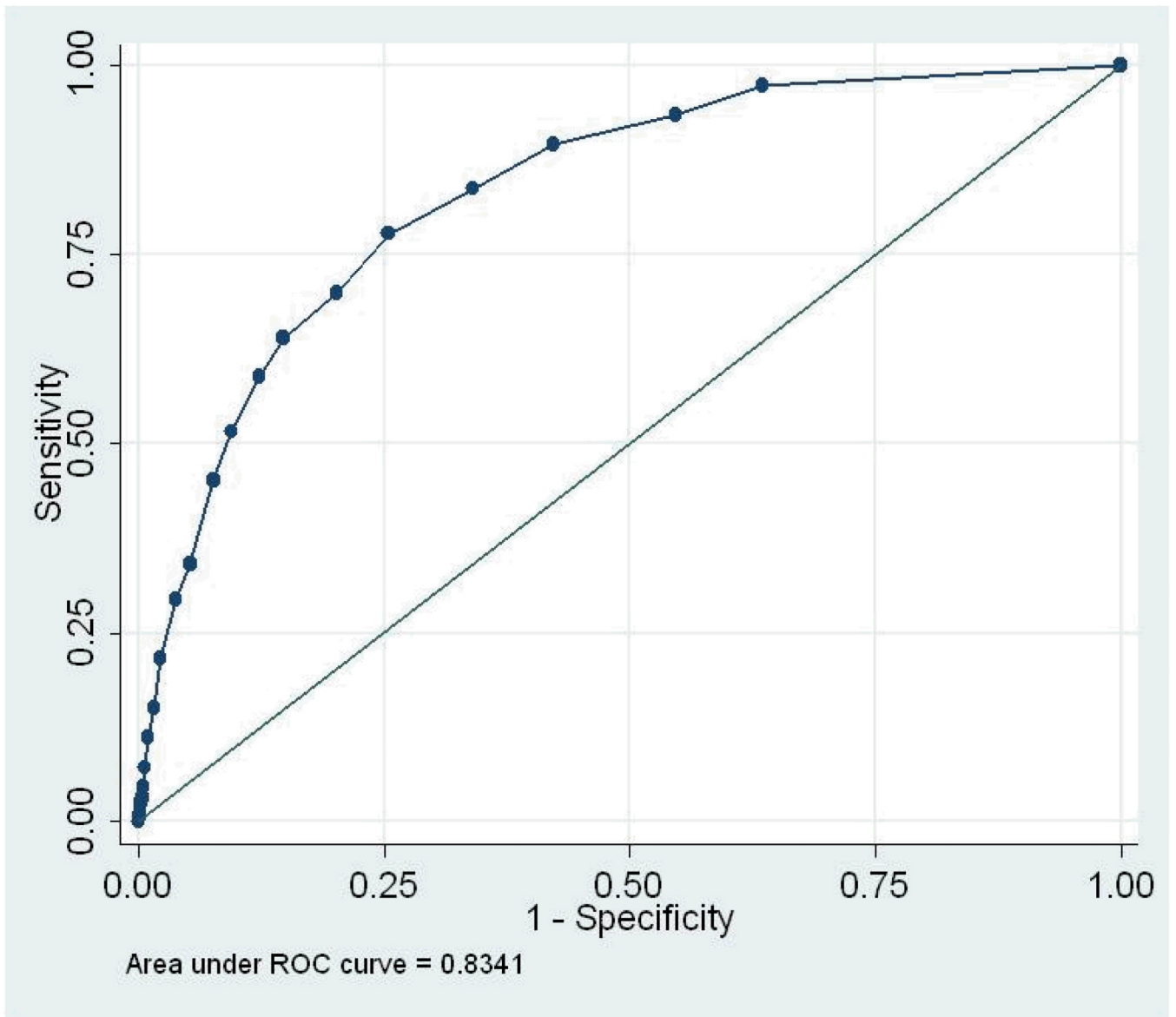


Figure 1.
Receiver Operating Characteristics Curve, any depressive disorder

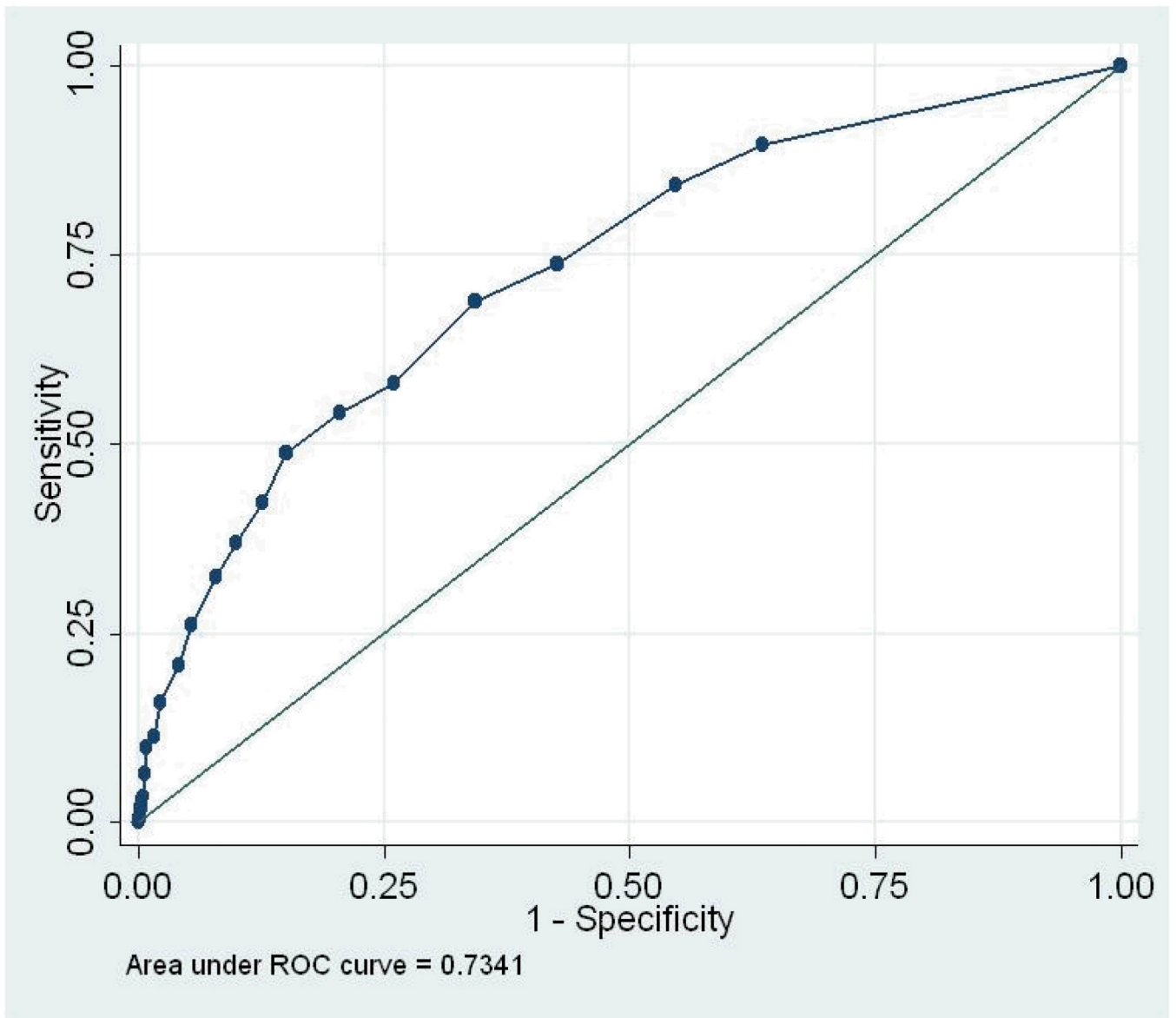


Figure 2.
Receiver Operating Characteristics Curve, any anxiety disorder

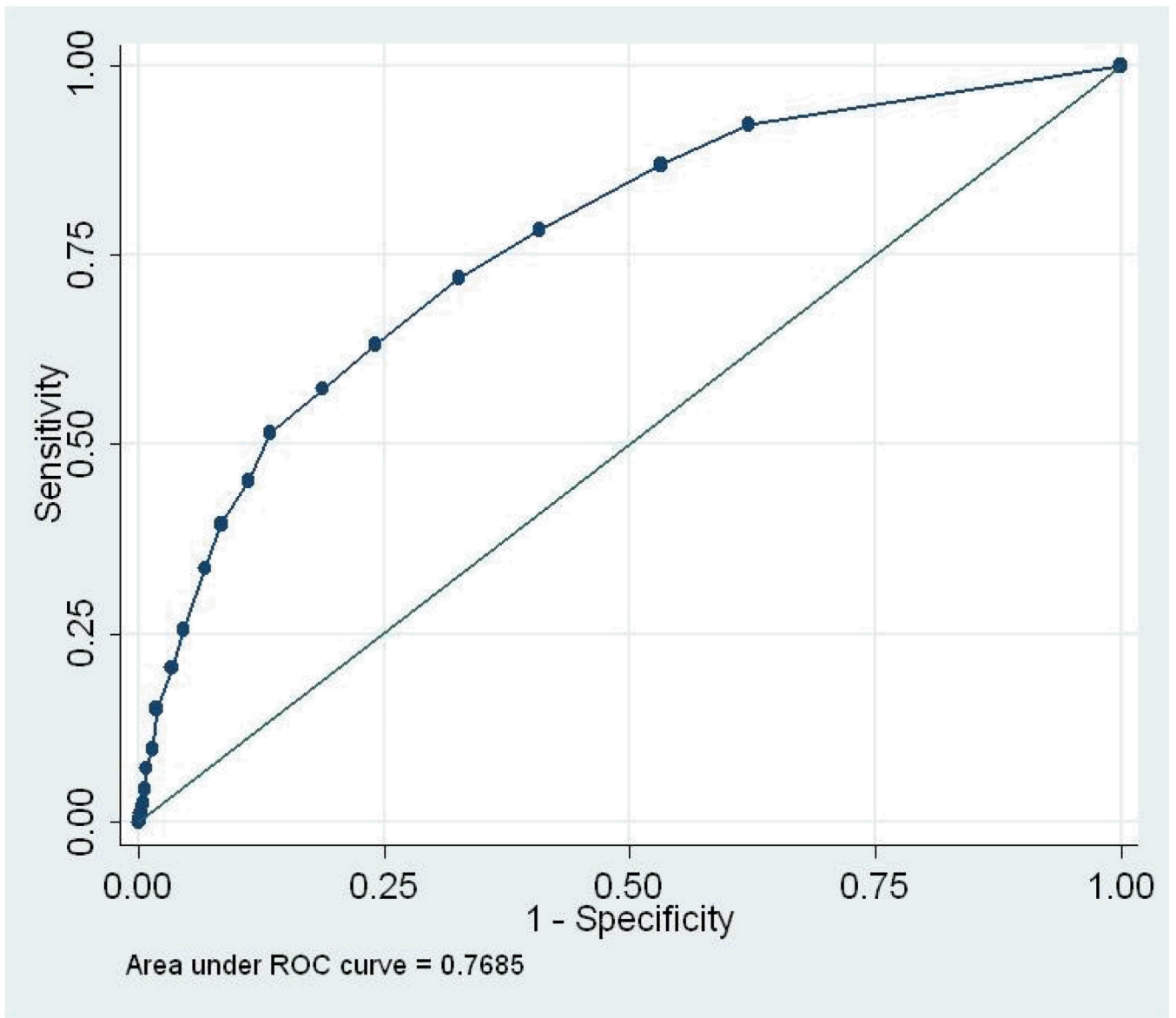


Figure 3.
Receiver Operating Characteristics Curve, any emotional disorder

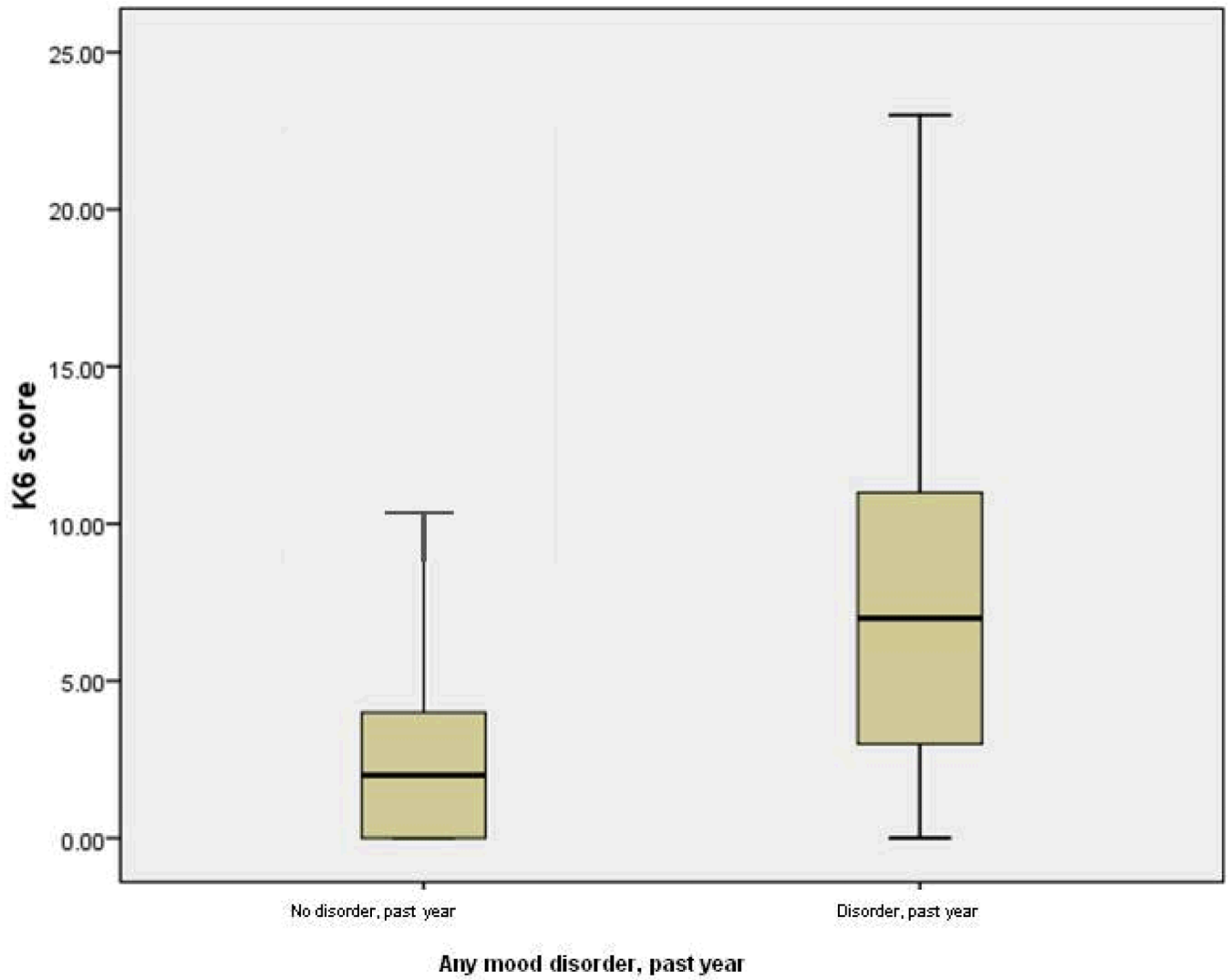


Figure 4.
Overlap and heterogeneity of K6 scores by diagnostic group

Table 1

Characteristics of the AI-SUPERPFP Samples (weighted)

	SOUTHWEST			NORTHERN PLAINS						
	Males (SM) n = 617	Females (SF) n = 829	Females (NF) n = 848	Males (SM) n = 790	Females (NF) n = 848	Females (NF) n = 848				
	%	99% CI ^a	%	99% CI ^a	%	99% CI ^a				
DEMOGRAPHICS										
Age										
15-24	25.7	23.8-27.8	NM	23.6	21.9-25.4	22.1	20.6-23.6	SM,NF	26.3	24.8-27.8
25-34	26.2	23.3-29.2		26.6	24.2-29.2	29.6	27.1-32.2		29.0	26.5-31.6
35-44	25.9	22.9-29.1		29.6	27.0-32.3	29.9	27.3-32.6		25.4	23.0-28.0
45+	22.2	20.1-24.5	NM	20.2	18.4-22.2	18.4	16.9-20.0	SM	19.3	17.9-20.8
Education										
Less than 12 years	29.2	24.7-34.2		27.4	23.5-31.7	24.8	21.1-28.9		27.9	24.1-32.1
HS Grad or GED	46.5	41.2-51.8		38.8	34.5-43.4	NM	48.9-58.3	SF, NF	41.3	36.8-46.0
Post-secondary	24.3	20.1-29.1	SF	33.8	29.6-38.2	SM, NM	17.9-25.8	SF, NF	30.8	26.6-35.3
Working for pay	62.5	57.4-67.4	NF	58.9	54.5-63.2		57.9-67.1	NF	50.1	45.5-54.6
Married or living as married	57.5	52.4-62.4		62.2	57.8-66.4	NM	44.2-53.8	SF	53.7	49.1-58.2
LIFETIME DSM-IV DISORDER PREVALENCE										
Depressive Disorder	8.9	6.3-12.5		13.0	10.2-16.4	NM	4.8-9.7	SF	9.3	6.9-12.5
Major Depressive Episode	8.5	6.0-12.1		12.3	9.6-15.7	NM	4.6-9.4	SF	9.1	6.7-12.2
Dysthymic Disorder	3.0	1.6-5.5		3.9	2.5-6.1		.8-3.3		3.0	1.8-5.0
Anxiety Disorder	14.4	11.1-18.6	SF, NF	22.6	19.0-26.8	SM, NM	8.2-14.2	SF, NF	20.9	17.4-25.0
Generalized Anxiety Disorder	2.4	1.2-4.7		4.1	2.6-6.3		.7-3.4		1.8	1.0-3.5
Panic Disorder	3.6	2.1-6.2		5.2	3.5-7.6		.8-3.5		3.1	1.8-5.2
Posttraumatic Stress Disorder	11.7	8.6-15.6	SF, NF	19.5	15.9-23.6	SM, NM	6.5-12.0	SF, NF	19.2	15.8-23.3
Substance Use Disorder	42.9	37.7-48.4	SF, NF	14.8	11.8-18.5	SM, NM, NF	38.3-48.1	SF, NF	31.0	26.7-35.6
Alcohol Abuse or Dependence	38.7	33.6-44.1	SF, NF	12.2	9.4-15.7	SM, NM, NF	36.3-46.0	SF, NF	28.8	24.6-33.3
Drug Abuse or Dependence	14.0	10.6-18.2	SF	5.2	3.5-7.8	SM, NM, NF	12.2-19.3	SF	10.9	8.2-14.3

^aCI = confidence interval

^bSuperscripts denote subgroups that differ significantly (p < .05)

Table 2
 Past-year Disorders: Area under the Curve (AUC) from Receiver Operating Characteristics (ROC) Curve

	overall			gender				tribe			
	AUC	95% CI ^a		male		female		SW		NP	
		AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI
depression/dysthymia	0.83	0.80, 0.87	0.88	0.84, 0.92	0.80	0.76, 0.85	0.82	0.77, 0.86	0.85	0.80, 0.89	
any anxiety disorder	0.73	0.70, 0.77	0.76	0.70, 0.82	0.72	0.67, 0.76	0.72	0.67, 0.78	0.75	0.70, 0.80	
any mood disorder	0.77	0.74, 0.80	0.81	0.76, 0.85	0.74	0.71, 0.78	0.76	0.72, 0.80	0.78	0.74, 0.81	

^a CI = confidence interval

Table 3
Regression of K6 Scores and Diagnoses on Physical Component Scale and Mental Component Scale

	Model 1		Model 2		Model 3	
	standardized		standardized		standardized	
	PCS ^a	MCS	PCS	MCS	PCS	MCS
age	-0.25 *	0.00	-0.25 *	0.01	-0.19 *	0.02
female	-0.03	-0.02	-0.03	-0.02	-0.01	-0.03
NP tribe	0.02	0.05 *	0.02	0.05 *	0.03	0.05 *
gender × tribe	-0.04	-0.02	-0.04	0.05	-0.03	-0.01
any emotional disorder, past year	-0.07 *	-0.19 *	-0.08 *	-0.20 *	-0.06 *	-0.20 *
K6 summary score	-0.19 *	-0.49 *	-0.18 *	-0.48 *	-0.16 *	-0.47 *
any emotional disorder, lifetime not past year			-0.06 *	-0.08 *	-0.03	-0.07 *
alcohol use disorder, past year					0.00	-0.04 *
alcohol use disorder, lifetime but not past year					-0.01	-0.02
physical diagnosis, past year					-0.25 *	-0.04 *
physical diagnosis, lifetime but not past year					-0.03	-0.01

^a PCS = Physical Component Scale; MCS = Mental Component Scale; higher score = greater quality of life

* $p < .05$