2004 NATIONAL SURVEY ON DRUG USE AND HEALTH

SERIOUS PSYCHOLOGICAL DISTRESS REPORT

Prepared for the 2004 Methodological Resource Book

Contract No. 283-03-9028
RTI Project No. 0208726.187.013

Authors:

Jeremy Aldworth
James R. Chromy
Misty Foster
Dave Heller
Scott Novak

Project Director: Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, MD 20857

Prepared by:

RTI International
Research Triangle Park, NC 27709

October 2005

2004 NATIONAL SURVEY ON DRUG USE AND HEALTH

SERIOUS PSYCHOLOGICAL DISTRESS REPORT

Prepared for the 2004 Methodological Resource Book

Contract No. 283-03-9028
RTI Project No. 0208726.187.013

Authors:

Jeremy Aldworth
James R. Chromy
Misty Foster
Dave Heller
Scott Novak

Project Director:
Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, MD 20857

Prepared by:

RTI International
Research Triangle Park, NC 27709

October 2005

Acknowledgments

This publication was developed for the Substance Abuse and Mental Health Services Administration (SAMHSA), Office of Applied Studies (OAS), by RTI International (a trade name of Research Triangle Institute), Research Triangle Park, North Carolina, under Contract No. 283-03-9028. Significant contributors at RTI include Jeremy Aldworth, James R. Chromy, Misty Foster, David Heller, and Scott Novak. Joyce Clay-Brooks word processed and formatted the report, Claudia Clark copyedited the report, and Thomas G. Virag is the RTI Project Director.

Table of Contents

Chapter **Page**

List of Figures

List of Tables

1. Introduction

For the 2004 National Survey on Drug Use and Health (NSDUH), serious psychological distress (SPD) was measured using the K6 screening instrument for nonspecific psychological distress (Furukawa, Kessler, Slade, & Andrews, 2003; Kessler et al., 2003). In previous NSDUH reports, the K6 scale was used to measure serious mental illness (SMI), but this was changed in 2004, and the reasons for this change are discussed in this report.

In the 2003 survey, the SMI module consisted of a broad array of mental health questions that preceded the K6 items. In the 2004 survey, the sample was split evenly between the "long-form" module used in the 2003 survey and a "short-form" module consisting only of the K6 items. Results from the 2004 survey showed that large differences in SPD prevalence rates occurred between the two modules, especially in the 18 to 25 age group. An attempt was made to adjust short-form estimates to match those from the long form, and these adjustments were used in State-level, but not national, estimates.

This report is organized as follows: Chapter 2 provides background information regarding SMI, the K6 instrument, and the reasons for the change from measuring SMI to measuring SPD; Chapter 3 discusses the construction of propensity strata required for psychometric analyses and adjustment procedures; Chapter 4 investigates psychometric properties of the K6; Chapter 5 discusses an adjustment method based on the individual K6 items; Chapter 6 discusses a gross adjustment method; and, finally, conclusions and a discussion on the application of adjusted values to estimates for the 2004 survey are given in Chapter 7.

2. Background

The official definition provided by the Substance Abuse and Mental Health Services Administration (SAMHSA) of adults with serious mental illness (SMI), based on a notice published in the *Federal Register* (SAMHSA, Center for Mental Health Services, 1993), is as follows:

> Pursuant to section 1912(c) of the Public Health Service Act, adults with serious mental illness (SMI) are persons: (1) age 18 and over and (2) who currently have, or at any time during the past year, had a diagnosable mental, behavioral, or emotional disorder of sufficient duration to meet diagnostic criteria specified within DSM-IV or their ICD-9-CM equivalent (and subsequent revisions) with the exception of DSM-IV "V" codes, substance use disorders, and developmental disorders, which are excluded, unless they co-occur with another diagnosable serious mental illness (3) that has resulted in functional impairment which substantially interferes with or limits one or more major life activities.

In National Survey on Drug Use and Health (NSDUH) reports prior to 2004, the K6 scale was used to measure SMI according to the above definition.[1] The K6 consists of six questions that ask respondents how frequently they experienced symptoms of psychological distress during the 1 month in the past year when they were at their worst emotionally. The use of this scale for SMI was based on a methodological study designed to evaluate several screening scales for measuring SMI in NSDUH. These scales consisted of a truncated version of the World Health Organization (WHO) Composite International Diagnostic Interview Short Form (CIDI-SF) scale (Kessler, Andrews, Mroczek, Üstün, & Wittchen, 1998), the K10/K6 scale of nonspecific psychological distress (Furukawa et al., 2003), and the WHO Disability Assessment Schedule (WHO-DAS) (Rehm et al., 1999).

The methodological study to evaluate the scales consisted of 155 respondents selected from a first-stage sample of 1,000 adults aged 18 or older. First-stage respondents were selected from the Boston metropolitan area and screened on the telephone to determine whether they had any emotional problems. Respondents reporting emotional problems at the first stage were oversampled when selecting the 155 respondents at the second stage. The selected respondents were interviewed by trained clinicians in respondents' homes using both the NSDUH methodology and a structured clinical interview. The first interview included the three scales described above using audio computer-assisted self-interviewing (ACASI). Respondents completed the ACASI portion of the interview without discussing their answers with the clinician. After completing the ACASI interview, respondents then were interviewed using the 12-month nonpatient version of the Structured Clinical Interview for DSM-IV (SCID; First, Spitzer, Gibbon, & Williams, 1997) and the Global Assessment of Functioning (GAF; Endicott, Spitzer, Fleiss, & Cohen, 1976) to classify respondents as either having or not having SMI. In the study, SMI was "operationally" defined as any 12-month DSM-IV disorder, other than a

[1] DSM-IV: *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. (American Psychiatric Association, 1994). ICD-9-CM: *International Classification of Diseases, Ninth Revision, Clinical Modification* (Central Office on ICD-9-CM, 1997).

substance use disorder, with a GAF score of less than 60, consistent with recommendations of a SAMHSA expert panel.

The data from the 155 respondents were analyzed using logistic regression analysis to predict SMI from the scores on the screening questions. Analysis of the model fit indicated that each of the scales alone and in combination were significant predictors of SMI and the best-fitting models contained either the CIDI-SF or the K10/K6 alone. Receiver operating characteristic (ROC) curve analysis was used to evaluate the precision of the scales to discriminate between respondents with and without SMI. This analysis indicated that the K6 was the most efficient screener. The results of the methodological study and the K10/K6 scale of nonspecific psychological distress are described in more detail in Kessler et al. (2003).

To create a score, the six items (DSNERV1, DSHOPE, DSFIDG, DSNOCHR, DSEFFORT, and DSDOWN) on the K6 scales were coded from 0 to 4 so that "all of the time" was coded 4, "most of the time" 3, "some of the time" 2, "a little of the time" 1, and "none of the time" 0, with "don't know" (DK) and "refuse" (REF) also coded 0. Summing across the transformed responses resulted in a score with a range from 0 to 24. Respondents with a total score of 13 or greater were classified as having a past year SMI. This cutpoint was chosen to equalize false positives and false negatives. At this cutpoint, sensitivity (SE) was 0.36 (0.08), specificity was 0.96 (0.02), and total classification accuracy was 0.92 (0.02) (Kessler et al., 2003).

The six questions comprising the K6 scale are given as follows:

DSNERV1 Most people have periods when they are not at their best emotionally. Think of 1 month in the past 12 months when you were the most depressed, anxious, or emotionally stressed. If there was no month like this, think of a typical month.

During that month, how often did you feel nervous?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

Response categories are the same for the following questions:

DSHOPE During that same month when you were at your worst emotionally . . . how often did you feel hopeless?

DSFIDG During that same month when you were at your worst emotionally . . . how often did you feel restless or fidgety?

DSNOCHR During that same month when you were at your worst emotionally . . . how often did you feel so sad or depressed that nothing could cheer you up?

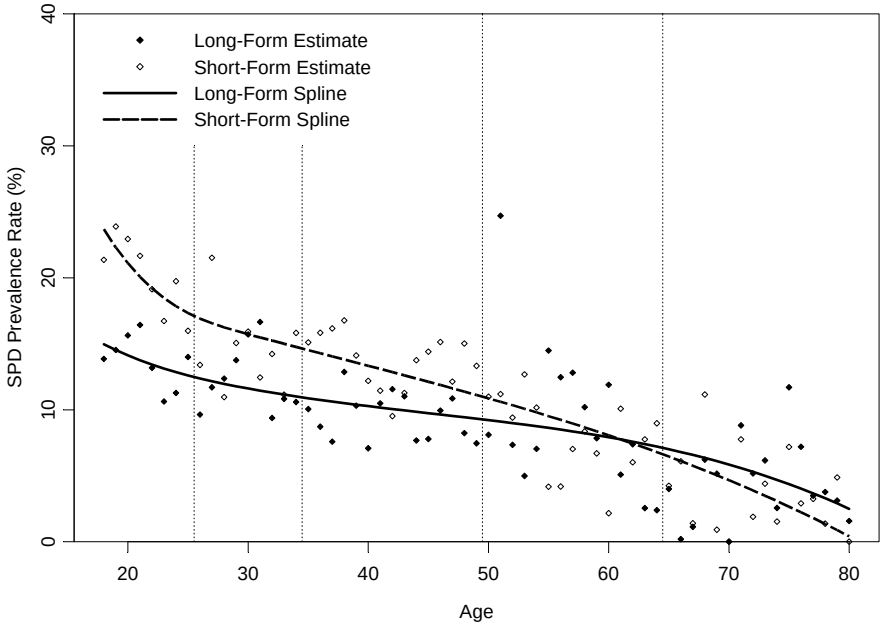
DSEFFORT During that same month when you were at your worst emotionally . . . how often did you feel that everything was an effort?

DSDOWN During that same month when you were at your worst emotionally . . . how often did you feel down on yourself, no good, or worthless?

In the 2003 survey, the mental health module contained a truncated version of the CIDI-SF scale, the K10/K6 scale, and the WHO-DAS scale, to mirror the questions used by Kessler et al. (2003). Thus, the module contained a broad array of questions about mental health (i.e., panic attacks, depression, mania, phobias, generalized anxiety, posttraumatic stress disorder, and use of mental health services) that preceded the K6 items, and the four extra questions in the K10 scale were interspersed among the items in the K6 scale. In the 2004 survey, the sample of respondents 18 or older was split evenly between the "long-form" module, which included all items in the mental health module used in the 2003 survey (Sample A), and a "short-form" module consisting only of the K6 items (Sample B). The "short-form" version was introduced to reduce interview time, removing questions that were not needed for estimation of SMI, and to provide space for a new module on depression. Inclusion of the "long-form" version in half of the sample was to measure the impact on the K6 responses of changing the context of the K6.

Results from the 2004 survey showed large differences between the two samples in both the K6 total score and the proportion of respondents with a K6 total score of 13 or greater. These differences were most pronounced in the 18 to 25 age group (see Figure 1). These differences suggest that the K6 scale is not context independent, that is, respondents appear to respond to the K6 items differently depending on whether or not the scale is preceded by a broad array of other mental health questions.

Figure 1. Cubic Splines Fitted to Age for Unweighted SPD Prevalence Estimates



Note: Long-form estimates are in black, and short-form estimates are in red.
Source: SAMHSA, Office of Applied Studies, 2004.

This raised concerns about the usefulness of the K6 scale in measuring SMI. There were other concerns as well. For example, the "face validity" of the K6 scale suggests that it may be more useful as a measure of psychological distress or affective-mood and anxiety-type disorders. Another concern is that the GAF criterion in the operational definition of SMI used by Kessler et al. (2003) might not best represent the definition in the *Federal Register* (SAMHSA, Center for Mental Health Services, 1993). A GAF score of less than 60 indicates moderate or worse functional impairment, which includes less severe cases than those implied by the definition in the *Federal Register*, which states that functional impairment should substantially interfere with or limit one or more major life activities. For this reason, SAMHSA decided to change the GAF criterion in the operational definition of SMI to reflect a GAF score of less than or equal to 50 (i.e., to indicate serious or worse functional impairment). A direct consequence of these concerns is that a decision was made that the K6 would no longer be used to measure SMI. However, the K6 data are still useful as an indicator of psychological distress. Therefore, using the same cutoff of 13, a new measure, serious psychological distress (SPD), is presented in 2004 NSDUH reports. Note that SPD potentially reflects a larger constellation of individuals than one might find using a stricter application of the *Federal Register* definition of SMI.

Further research is under way in the Office of Applied Studies (OAS) to try to develop a valid module that will provide SMI estimates efficiently (i.e., with a small number of questions) from the survey. Unfortunately, given the apparent context effects and data collection differences, it may not be feasible to use results from studies that are done to develop and validate scales (including K6) for other surveys.

3. Propensity Strata

Since approximately one half of the questionnaires in the 2004 survey contained the short-form serious psychological distress (SPD) module, an attempt was made to "salvage" the K6 scores obtained from those questionnaires by adjusting them to match the K6 scores obtained from questionnaires containing the long-form module. The direction of this adjustment was based on the fact that the long-form module was the same as the one used by Kessler et al. (2003).

To begin this adjustment process, we first define *propensity* as the difference in SPD prevalence rates between short and long forms. The propensity is therefore a measure of the bias due to the short form, assuming that the long form is the gold standard.

Figure 1 shows that propensity varies by age, and there may be other covariates within whose levels the propensity also varies. We therefore need to construct a model describing how propensity differs across (i.e., interacts with) different covariates. Once a model has been constructed, we then need to create propensity strata in such a way that each stratum contains roughly similar propensity estimates. For this purpose, a weighted logistic regression model was developed with the SPD indicator variable (i.e., SPD = 1 if SPD score $\geq$ 13; and is 0 otherwise) as the dependent variable, and the independent variables consisted of the sample variable (i.e., indicating long form or short form), several demographic and drug use covariates, and some interactions with the sample variable. In the model decided upon, the Sample * Age * Race/Ethnicity interaction was statistically significant. Therefore, based on this model, the propensities and corresponding standard errors were estimated for each cell in the two-way table whose dimensions were age group (18 to 25, 26 to 34, 35 to 49, 50 to 64, and 65 or older) and race/ethnicity (white, black or African American, other, and Hispanic or Latino). Based on the resulting propensities, five strata were constructed as follows, where the two symbols identifying the race/ethnicity and age group of each element in the strata are self-explanatory (e.g., "W1" refers to whites aged 18 to 25):

- Stratum 1 (propensity < -5.00%): H4, H5
- Stratum 2 (-5.00% ≤ propensity < -1.67%): W4
- Stratum 3 (-1.67% ≤ propensity < 1.67%): W2, W5, O2
- Stratum 4 (1.67% ≤ propensity < 5.00%): W3, B3, B4, B5, O1, O5, H2, H3
- Stratum 5 (5.00% ≤ propensity): W1, B1, B2, O3, O4, H1

The construction of these strata is somewhat arbitrary (i.e., in terms of the number of strata and the designation of boundaries), and since some of the propensity estimates have low precision (e.g., for those groups that have small sample sizes, such as H5 (elderly Hispanics or Latinos), there is likely to be some classification error in assigning estimates to strata.

4. Psychometric Properties of the K6 Items

This section describes an investigation in the psychometric properties of the K6 instrument and an evaluation of potential discrepancies between the 2004 long (Sample A) and short (Sample B) forms. First, we calculated an alpha coefficient to examine internal consistency of the six items in the serious psychological distress (SPD) module (Cronbach, 1951). The formula for this coefficient is as follows:

$$R_{xx} = (n/n-1)\left(S^2 - \sum S_i^2 / S^2\right),$$

where n = number of items (6), S^2 is the variance of the summated scores, and $\sum S_i^2$ is the sum of the variances of the items comprising the scale. Typically, a scale alpha greater than 0.70 has been suggested as indicating acceptable internal consistency (Nunnally, 1978). As noted in Table 1, the SPD scale had desirable internal consistency for the entire sample and for each subsample (alpha > 0.86).

Table 1. Psychometric Properties of K6 SPD Module

Sample	Alpha	Standardized Factor Loadings					
		DSNERV1	DSHOPE	DSFIDG	DSNOCHR	DSEFFORT	DSDOWN
All	0.912	0.686	0.836	0.738	0.853	0.777	0.815
Short	0.944	0.736*	0.841	0.776*	0.858	0.784	0.822
Long	0.863	0.586*	0.826	0.667*	0.844	0.759	0.796
White, 18-25	0.857	0.627	0.862	0.696	0.845	0.776	0.850
Short	0.930	0.674*	0.868	0.732*	0.863	0.784	0.863
Long	0.862	0.540*	0.854	0.634*	0.817	0.761	0.828
Black or African American, 18-25	0.838	0.654	0.848	0.730	0.816	0.588	0.817
Short	0.929	0.712*	0.850	0.759	0.842	0.641*	0.836
Long	0.798	0.535*	0.851	0.676	0.770	0.487*	0.783

* Denotes statistically significant differences between long and short forms ($p < 0.05$).

Next we examined the factor structure of the K6 scale to examine the degree to which individual items accounted for variation in SPD scores. This process was confirmatory given that we assumed that the scale has only a single dimension. Thus, we examined whether the items shared a common conceptual meaning in the sense that they were associated with the same general construct—serious psychological distress.

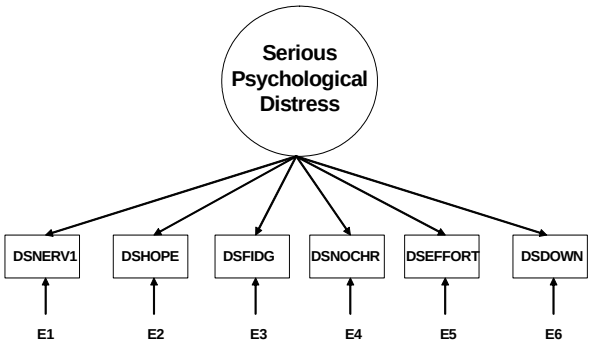
A brief note is warranted about the nature of the scaled items in relation to the general construct. Scales are often referred to as latent in the sense that they cannot be observed directly. Rather, they are defined in reference to observed variables that are assumed to be associated with the construct. There are some assumptions that need to be made about the underlying causal structure between the observed variables and the latent construct. One possible assumption is that the underlying factor "causes" each individual item. An example is that serious psychological

distress cannot be observed directly but influences the presence and magnitude of its indicators, such as nervousness. Therefore, SPD influences each response to the six items in the SPD module (see Figure 2).

This can be translated to a regression framework in that each item *j* for person *i* is influenced by the overall mean ($\beta 0$), slope of the latent factor ($\beta 1$), and a residual error term (E_{ij}). Thus,

$$Y_{ij} = \beta 0_i + \beta 1_i(SPD_{ij}) + (E_{ij}).$$

Figure 2. Hypothesized Causal Structure of SPD Module



In this approach, each $\beta 1$ slope is typically referred to as a either a weight or factor loading. By standardizing these slopes (range 0-1), we can assess the amount of variation in each item that is attributable to the overall latent factor. Moreover, statistical tests can be estimated through nested comparisons in which the factor loadings are allowed to vary between samples. All analyses were conducted using Mplus (http://www.statmodel.com), adjusting for features of the complex sampling design and weighting scheme within the context of continuous indicator variables in the measurement model.

Table 1 presents the standardized factor loadings for each item among the entire sample and separately by instrument (e.g., long and short forms). Among the entire sample, DSNERV1 was more significantly related to SPD for those receiving the short form compared with the long form. This relationship held across groups and, specifically, the two groups with the highest propensity for reporting bias, white and black or African-American young adults (aged 18 to 25). Interestingly, differences in the factor loadings for DSFIDG were present for the entire sample and, specifically, white young adults. No such effect was present for black or African-American young adults. However, differences were observed for DSEFFORT, where those who answered the short form placed greater weight in answering this item compared with those who were administered the long form.

The pattern of the factor loadings suggests that the long form tended to have more measurement error in the individual items compared with the short form. This is not to say that the short form is a better measure of SPD. Rather, respondents appear to have used information

from all available questions to form their response set compared with the long-form respondents, who most likely used information from the preceding psychopathology questions to frame their responses.

5. Model-Based Item Adjustment

After deriving the demographic groups with the highest propensity for response bias (see Chapter 3), we then sought to identify candidate items in the K6 that may have accounted for differential response. As noted earlier, propensity was calculated as the difference in observed serious psychological distress (SPD) probability between the short (Sample B) and long (Sample A) forms. Chapter 4 described how we examined the performance of the items in reference to the latent construct of SPD. In this chapter, we describe how we employed a direct statistical test to identify candidate items that significantly differed on the observed distribution between samples within each propensity stratum. Items in the short form that were significantly different from those in the long form were assumed to be unreliable and biased.

After candidate items were identified, the next step in the analytic plan involved imputing these items in the short-form sample from responses in the long form. As these items were assumed to be unreliable and biased, we treated this situation as a generalized case of data that were missing at random (MAR). This formulation posits that the pattern of missingness for each item can be predicted by observed covariates. Therefore, we could predict an individual response for each item from the non-K6 items. It should be noted that this assumption is distinguished from data that are missing completely at random (MCAR), which presumes that the pattern of missingness is arbitrary and is not dependent on observed or unobserved covariates. Given our assumption of MAR, the resulting predicted estimates were unbiased and had minimum variance because they were based on likelihoods. After formulating a model among respondents in Sample A that maximized the prediction of each item, the resulting parameter estimates were exported to Sample B to output the predicted response for each candidate item. A new SPD score was created for Sample B such that it represented the summation of the observed and imputed K6 items.

Within the highest propensity stratum (see Tables 2 and 3), the likelihood ratio tests, which adjusted for the design and weighting characteristics of the survey, revealed that all items were significantly different ($p < 0.05$) between the long and short forms. Given the pattern of significance across all items within the high propensity groupings, the imputation scheme as specified presents a problem as the data in Sample B (short form) would be treated as massively multivariate missing data. Thus, the assumptions of model-based imputation methods become more difficult to satisfy because the functional form would be more complex and thus unlikely to hold between years of survey administration. Classical psychometric assumptions hold that item imputation is tenable when less than 20 percent of the items in a scale are missing (Healy & Westmacott, 1956; Little & Rubin, 1987). In the case of the SPD module, 100 percent of the items would need to be treated as missing—a rather untenable strategy. Overall, gross adjustment methods, which are based on less stringent assumptions, are warranted.

Table 2. Likelihood Ratio Chi-Squared Tests for SPD Items Within Strata, by Sample

Variable	All	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5
DSNERV1	Sig	ns	< 0.05	< 0.001	< 0.001	< 0.001
DSHOPE	Sig	ns	ns	< 0.001	< 0.001	< 0.001
DSFIDG	Sig	ns	< 0.05	< 0.001	< 0.001	< 0.001
DSNOCHR	Sig	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
DSEFFORT	Sig	ns	ns	< 0.001	< 0.001	< 0.001
DSDOWN	Sig	ns	ns	< 0.001	< 0.001	< 0.001

Note: "ns" denotes $p > 0.05$.

Table 3. Likelihood Ratio Chi-Squared Tests for SPD Items Within Stratum 5, by Sample

Variable	All	White (18-25)	Black or African American (18-25)	Black or African American (26-34)	Other (35-49)	Other (50-64)	Hispanic or Latino (18-25)
DSNERV1	Sig	< 0.001	< 0.001	< 0.05	ns	ns	< 0.001
DSHOPE	Sig	< 0.001	< 0.001	ns	< 0.05	ns	< 0.001
DSFIDG	Sig	< 0.001	0.021	ns	< 0.05	ns	< 0.001
DSNOCHR	Sig	< 0.001	ns	ns	ns	ns	< 0.001
DSEFFORT	Sig	< 0.001	0.001	ns	< 0.05	ns	< 0.001
DSDOWN	Sig	< 0.001	0.001	ns	< 0.05	< 0.05	< 0.05

Note: "ns" denotes $p > 0.05$.

6. Adjustments Based on the CDF Method

The cumulative distribution function (CDF) adjustment method transforms the distribution of one variable so that it matches the distribution of another variable. This approach is commonly used in digital image processing, where it also is referred to as histogram specification (e.g., Jain, 1989).

Consider two random variables X and Y. Suppose we require Y to be transformed in order to match the distributional properties of X. A CDF transformation can achieve this exactly if X and Y are continuous, or approximately if X and Y are discrete (Jain, 1989, p. 224). A modification of this method that includes an extra randomization step has been applied to the discrete serious psychological distress (SPD) score variable, and as a consequence, provides an exact mapping in the expectation of that randomization.

Suppose X and Y are *continuous* random variables with CDFs $F(x) = P(X \le x)$ and $G(y) = P(Y \le y)$, respectively. It is easy to show that $U = F(X)$ and $V = G(Y)$ both possess Uniform (0,1) distributions. Because $F(X) \stackrel{d}{=} G(Y)$ and F^{-1} is a well-defined function over the support of X, $X \stackrel{d}{=} F^{-1}(G(Y))$, where $\stackrel{d}{=}$ denotes equality in distribution. From this we may obtain a suitably CDF-transformed Y, namely Y^*, defined as

$$Y^* = F^{-1}(G(Y)) \stackrel{d}{=} X, \text{ as desired.}$$

Now suppose X and Y are *discrete* random variables, with CDFs $F(x) = P(X \le x)$; $x = 0,1,...,24$ and $G(y) = P(Y \le y)$; $y = 0,1,...,24$, respectively. Because discrete CDFs are step functions, there may not exist an x such that $G(y_i) = F(x)$, but we can say that $y_i \le x_j \Rightarrow G(y_i) \le F(x_j)$ and $y_i \ge x_j \Rightarrow G(y_i) \ge F(x_j)$. Therefore, we can define a transformed realization of Y in terms of three cases, as follows:

1. If $G(y_i) < F(0)$ then $y_i^* = 0$
2. If $G(y_i) > F(24)$ then $y_i^* = 24$
3. If $F(x_{j-1}) \le G(y_i) \le F(x_j)$ then $y_i^* = \begin{cases} x_j, \text{ with probability } \frac{G(y_i)-F(x_{j-1})}{F(x_j)-F(x_{j-1})} \\ x_{j-1}, \text{ with probability } \frac{F(x_j)-G(y_i)}{F(x_j)-F(x_{j-1})} \end{cases}$

Consider an example of how this would operate for the nontrivial third case. Suppose X and Y represent long- and short-form SPD scores and that we obtain $G(14) = 0.879$, $F(12) = 0.853$, and $F(13) = 0.886$. Because $F(12) < G(14) < F(13)$, a score of 14 from the short form will be adjusted to a value of either 12 or 13 as follows: Generate a realization r from a Uniform (0,1) distribution, and, if $r < (0.879 - 0.853)/(0.886 - 0.853) = 0.788$, adjust that value to 13; otherwise, adjust it to 12. Repeat for each value of 14 from the short form. Then perform this process for every other SPD score from the short form. The CDF of the transformed short-form

SPD scores will be identical to that of the long-form SPD scores in the expectation of the Uniform (0,1) randomization process.

The CDF adjustment method was applied within each of the five strata defined in Chapter 2. The effects of this adjustment to the marginal SPD prevalence estimates of several demographic and drug use variables can be seen in Table 4. Note that of the 24 comparisons between marginal estimates based on unadjusted short- and long-form SPD scores, 18 are statistically significant. There are no statistically significant differences between marginal estimates based on adjusted short- and long-form scores.

Table 4. Comparison of SPD Prevalence Rates Based on Unadjusted and Adjusted Short- and Long-Form Scores

Category	Short Form	Adjusted Short Form	Long Form
Total	12.18[b]	10.22	9.95
18-25	20.21[b]	13.42	13.71
26-34	14.50[a]	12.37	12.10
35-49	13.69[b]	9.89	9.48
50-64	8.23	10.78	10.04
65+	5.08	4.95	5.03
Male	9.39[a]	7.62	7.70
Female	14.78[b]	12.64	12.03
Not Hispanic or Latino	12.18[b]	10.12	9.82
White	12.22[b]	10.66	10.25
Black or African American	11.92[b]	7.60	8.11
Other	12.18[a]	8.22	7.98
Hispanic or Latino	12.20	10.98	10.81
Northeast	11.67	9.73	9.71
Midwest	12.38[b]	10.20	10.13
South	12.37[b]	10.04	9.61
West	12.12	10.96	10.50
Large Metro	11.94[b]	9.70	8.86
Small Metro	12.78	11.10	12.00
Rural	11.93[a]	10.39	9.68
Marijuana	16.83[b]	13.80	13.59
Cocaine	19.95[a]	16.51	16.86
Cigarettes	13.63[b]	11.54	11.10
Alcohol	12.68[b]	10.60	10.23

[a] Contrast with long-form estimate is statistically significant ($p < 0.05$).
[b] Contrast with long-form estimate is statistically significant ($p < 0.01$).

7. Conclusions

As described in Chapter 3, a logistic regression model was used to estimate differences between short- and long-form serious psychological distress (SPD) prevalence rates (i.e., propensities). Several demographic and drug use covariates were included in the model, and it was found that the propensities varied according to race/ethnicity and age group. Five propensity strata based on race/ethnicity and age group were constructed from the results of this analysis. Chapter 4 presented the psychometric analysis of the six K6 items of the SPD, and Chapter 5 presented the model-based item adjustment method, both of which suggest that a gross adjustment approach might be more appropriate than an item-based adjustment approach. As a consequence, the cumulative distribution function (CDF) (gross) adjustment method was applied within the five propensity strata, and the method appeared to work quite well in adjusting the marginal estimates of a number of important demographic and drug use variables (see Table 4 in Chapter 6). This method also was shown to be fairly robust to the way the propensity strata were defined.

Consideration also was given to the use of the logistic regression model as a means to provide adjustments, in addition to providing propensity estimates. However, while this approach is useful for estimating propensities, it is not useful in determining how to adjust *individual* short-form respondents' SPD prevalence rates to match those of long-form respondents within covariate profiles. Using this approach, the only way to match prevalence rates would be to use long-form prevalence estimates in place of short-form prevalence estimates within covariate profiles. This is equivalent to *discarding* all short-form data after the logistic regression model has been fitted. A similar argument applies to the use of polytomous logistic regression models to estimate differences between short- and long-form SPD scores.

Before the CDF adjustment method was developed, consideration also was given to ad hoc CDF adjustments to differences between short- and long-form SPD scores within covariate profiles, estimated from, say, polytomous regression models. For example, if the average difference between short- and long-form SPD scores for a particular covariate profile (e.g., white females aged 12 to 17 in the West) was 1.7, then all short-form SPD scores would be reduced by that amount in the profile. However, there are a couple of problems with this ad hoc approach. First, this approach is equivalent to shifting the entire CDF of short-form scores to the left, creating a set of adjusted values ranging from -1.7 to 22.3 instead of 0 to 24. Second, while this approach might force SPD scores to match on average within a profile, there is no guarantee that they would match at the SPD cutpoint of 13, which defines prevalence rates. A variation to this approach would be to multiply short-form scores by a factor that forced the scores to match on average, but this is equivalent to rescaling the short-form CDF so that all scores are shrunk toward zero. Neither of these ad hoc methods was used.

Adjusted short-form SPD scores and prevalence rates (based on the CDF adjustment method) were not used to derive national estimates for the 2004 survey. National estimates used a much finer categorization for some of the demographic and substance use variables than were used in the analyses described in this report. For example, the second oldest age category in this report included all ages between 50 and 64, but in the national estimates, this category was subdivided further into the following categories: 50 to 54, 55 to 59, and 60 to 64. So, while the

adjusted values performed well at the level of categorization shown in Table 4, some notable discrepancies were observed between adjusted short-form and corresponding long-form prevalence rates at some of the finer categorizations, particularly those with small sample sizes. For this reason, national estimates of SPD scores and prevalence rates were derived from only long-form data.

Adjusted short-form SPD scores and prevalence rates were used to derive State-level estimates based on pooled 2003 and 2004 survey data. Since State-level estimates used a much coarser categorization of demographic and substance use variables than national estimates, the problem of discrepancies observed at the finer categorization of national estimates did not occur. In addition, unlike national estimates, which were based on large sample sizes, State-level estimates were typically based on small sample sizes. Hence, it was necessary to use all the data available, including the adjusted short-form data.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (DSM-IV)* (4th ed.). Washington, DC: American Psychiatric Association.

Central Office on ICD-9-CM. (1997). *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)*. Chicago, IL: American Hospital Association.

Cronbach, L. J. (1951). Coefficient Alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry, 33*(6), 766-771.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP)*. New York, NY: New York State Psychiatric Institute, Biometrics Research.

Furukawa, T. A., Kessler, R. C., 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. *Psychological Medicine, 33*(2), 357-362.

Healy, M. J. R., & Westmacott, M. (1956). Missing values in experiments analysed on automatic computers. *Applied Statistics, 5*, 203-206.

Jain, A. K. (1989). *Fundamentals of digital image processing*. Upper Saddle River, NJ: Prentice-Hall.

Kessler, R. C., Andrews, G., Mroczek, D., Üstün, T. B., & Wittchen, H.-U. (1998). The World Health Organization Composite International Diagnostic Interview Short Form (CIDI-SF). *International Journal of Methods in Psychiatric Research, 7*, 171-185.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S. L., Manderscheid, R. W. , Walters, E. E., & Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry, 60*(2), 184-189.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Rehm, J., Üstün, T. B., Saxena, S., Nelson, C. B., Chatterji, S., Ivis, F., & Adlaf, E. (1999). On the development and psychometric testing of the WHO screening instrument to assess disablement in the general population. *International Journal of Methods in Psychiatric Research, 8*, 110-123.

Substance Abuse and Mental Health Services Administration, Center for Mental Health Services. (1993, May 20). Final notice [Final definitions for: (1) Children with a serious emotional disturbance, and (2) adults with a serious mental illness]. *Federal Register, 58*(96), 29422-29425.