International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 19(Supplement 1): 88–98 (2010)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/mpr.315

Estimating prevalence of serious emotional disturbance in schools using a brief screening scale

FAN LI,[1] JENNIFER GREIF GREEN,[2] RONALD C. KESSLER[2] & ALAN M. ZASLAVSKY[2]

1 Department of Statistical Science, Duke University, Durham, NC, USA
2 Department of Health Care Policy, Harvard Medical School, Boston, MA, USA

Key words
K6 screening scale, psychiatric epidemiology, serious emotional disturbance, small-area estimation

Correspondence
Ronald C. Kessler, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston MA 02115, USA.
Telephone (+1) 617-432-3587
Fax (+1) 617-432-3588
Email: Kessler@hcp.med.harvard.edu

Received 8 March 2010; accepted 8 March 2010

Abstract

Information about the prevalence of serious mental illness (SMI) among adults or serious emotional disturbance (SED) among youth in small domains such as counties, states, or schools is valuable for mental health policy planning purposes, but prohibitively expensive to collect with semi-structured surveys. Commonly used synthetic estimation methods yield imprecise estimates. An improved method is described here that combines information about socio-demographic covariates with screening scale scores obtained from a sample of individuals, using a prediction equation derived from a Bayesian multilevel regression model with bivariate outcomes fitted to a larger population survey. This method is illustrated using K6 screening scale scores to predict school-level prevalence of SED in the sample of 282 schools that participated in the National Comorbidity Survey Replication Adolescent Supplement. Respondents completed a diagnostic interview that was used to define DSM-IV SED. SED prevalence varied significantly across schools and was strongly correlated with aggregate K6 scores ($\rho = 0.70$). Calculations suggest that near-maximum precision of school-level SED prevalence estimates could be attained with K6 samples of 200 students per school. This modeling approach holds great promise for generating accurate estimates of SMI/SED in small-area planning units based on K6 scores collected in ongoing health tracking surveys. *Copyright © 2010 John Wiley & Sons, Ltd.*

Introduction

Information about the prevalence of serious mental illness (SMI) among adults or serious emotional disturbance (SED) among youth in small areas such as counties or states can be valuable for mental health policy planning purposes. However, the costs of carrying out surveys to obtain direct estimates of SMI/SED in small areas are prohibitive. As a result, most efforts to estimate SMI/SED in small areas are based on relatively simple 'synthetic estimation' methods that reweight national surveys to match the joint distributions of socio-demographic variables in the small areas (Goldsmith *et al.*, 1998; Hudson, 2009; Kessler *et al.*, 2003; Thomas *et al.*, 2009). These synthetic estimates are usually imprecise because socio-demographic variables are not strongly associated with mental disorders (Kessler *et al.*, 1999).

A more promising approach to small-area estimation is to administer a brief screening scale for SMI/SED to a sample of respondents in each of the small areas of interest and to estimate the prevalence of the outcome from a regression equation that uses these screening scale scores, together with socio-demographic variables, to predict a gold-standard measure of the outcome. The potential for data of this sort exists at the county and state level in the USA because all three of the major US national health tracking surveys, the US National Health Interview Survey (www.cdc.gov/nchs/nhis.htm), the Centers for Disease Control Behavioral Risk Factors Surveillance Survey (www.cdc.gov/BRFSS), and the Substance Abuse and Mental Health Services Administration (SAMHSA) National Household Survey on Drug Use and Health (www.oas.samhsa.gov/nhsda.htm), have administered the K6 screening scale (Kessler *et al.*, 2002) and the combined sample size across the three surveys (close to 500,000 respondents per year) would be large enough to yield a meaningful probability sample of people in most counties in the USA. Even in areas where this combined sample is too small for analysis (e.g. in particular towns or low-income neighborhoods of cities, in particular schools), the K6 is so brief that a screening survey could be administered in these areas without great expense. Reports in this issue show that the K6 is a good predictor of SMI among adults (Kessler *et al.*, 2010) and of SED among youth (Green *et al.*, 2010). To estimate the prediction equation, it is also necessary to obtain a gold-standard assessment of the outcome linked to the screening scale in an appropriately designed sub-sample of respondents, not necessarily part of the same survey that provides small-area data with national coverage.

This approach can yield highly accurate estimates of the prevalence of mental disorders in small areas if the screening scale is as strongly correlated with the gold-standard measure of outcome as the K6 scale is with SMI and SED. We illustrate this by using scores on the K6 to predict the school-level prevalence of SED, basing our models on a sample of 282 schools that participated in the National Comorbidity Survey Replication Adolescent Supplement (NCS-A; Kessler *et al.*, 2009b). This example is important because schools are leading providers of mental health services to children and adolescents in the USA, with studies estimating that up to 80% of youths who receive mental-health-related services obtain them in schools (Burns *et al.*, 1995; Leaf *et al.*, 1996). Yet only half of all schools provide on-site mental health counseling services and there are considerable discrepancies in service availability related to characteristics such as region, urbanicity, and school size (Slade, 2003). In the public service sector, resource allocation is partially driven by distribution of need, defined as disorder severity and functional impairment. In particular, US Block Grant funds for community mental health services are allocated by the Center for Mental Health Services to provide services for children with SED, which is defined by the SAMHSA as the presence of a Diagnostic and Statistical Manual of Mental Disorders (DSM) diagnosable mental, behavioral, or emotional disorder that results in substantial impairment in the child's role or functioning in activities in their home, school, or community (Substance Abuse and Mental Health Services Administration, 1993, p. 29425). As a result, information about the school-level prevalence of SED across all schools could be of great value in targeting school-based treatment resources. However, estimates of this sort are not available. The methodology described here provides a practical way to obtain such data from school-based K6 screening scale surveys that could be completed in no more than 5 minutes.

Methods

Sample

The NCS-A was carried out between February 2001 and January 2004. Adolescents (ages 13–17 years) were interviewed face-to-face in dual-frame household and school samples (Kessler *et al.*, 2009a,b). As our focus is on school-level estimates, this report uses only the school sample, consisting of 9244 adolescents from a representative sample of 320 schools in the counties that participated in the National Comorbidity Survey Replication (NCS-R) (Kessler and Merikangas, 2004). The conditional (on school participation) adolescent response rate was 82.6%. Although the proportion of initially selected schools that participated was low (28.0%), non-participators were replaced with matched replacement schools. Comparison of household-sample respondents who attended non-participating schools with school sample respondents from replacement schools found no evidence of bias in estimates of either prevalence or correlates of disorders (Kessler *et al.*, 2009a), suggesting that the matched replacement procedures corrected for any biases that might otherwise have occurred as a result of the low school-level response rate. One parent or surrogate (henceforth referred to as parents) was asked to complete a self-administered questionnaire (SAQ) about the participating adolescent's developmental history and mental health. The SAQ conditional (on adolescent participation) response rate was 82.5% in the household sample and 83.7% in the school sample.

NCS-A schools were primarily public (86%) rather than private or parochial, and most were junior high

(28%) or high schools (46%) rather than those incorporating a larger range of grades (26%). On the Educational Climate Index, a measure of socio-economic status that incorporates educational aspects of social status, 22% were low, 38% were medium, 29% were medium–high, and 10% were high. Schools from which fewer than 10 students participated in the NCS-A were excluded from the sample, leaving 9022 students drawn from 282 schools (mean 32 students per school, inter-quartile range 28–37) in 42 geographically defined sampling strata.

Written informed consent was obtained from parents or guardians before approaching adolescents for participation in the NCS-A. Written adolescent assent was then obtained from adolescents before surveying either adolescents or parents. Each respondent was given $50 for participation. Recruitment–consent procedures were approved by the Human Subjects Committees of Harvard Medical School and the University of Michigan. Completed survey data were weighted for residual discrepancies between sample and population socio-demographic and geographic distributions (Kessler *et al.*, 2009a, b). The weighted composite sample socio-demographic distributions closely approximate those of the Census population (Kessler *et al.*, 2009a).

Measures

SED

Adolescents were administered a modified version of the Composite International Diagnostic Interview (CIDI), a fully structured interview designed for use by trained lay interviewers (Kessler and Üstün, 2004). The DSM-IV disorders assessed and considered here include mood disorders (major depressive disorder or dysthymia, bipolar I–II disorder), anxiety disorders (panic disorder with or without agoraphobia, agoraphobia without history of panic disorder, social phobia, specific phobia, generalized anxiety disorder, post-traumatic stress disorder, separation anxiety disorder), and behavior disorders (attention-deficit/hyperactivity disorder, oppositional-defiant disorder, conduct disorder, intermittent explosive disorder, and eating disorders such as anorexia nervosa, bulimia nervosa, binge-eating disorder). Parents provided diagnostic information about major depressive disorder/dysthymia, attention-deficit/hyperactivity disorder, oppositional-defiant disorder, and conduct disorder, the disorders for which parent reports have been shown to play the largest part in diagnosis. Parent and adolescent reports were combined at the symptom level using an 'or' rule (i.e. the symptom was considered present if endorsed by either respondent). All diagnoses applied DSM-IV organic exclusion rules and DSM-IV diagnostic hierarchy rules with the exception of oppositional-defiant disorder, which was defined with or without conduct disorder. Prevalence was assessed in multiple time frames, but only past 12-month prevalence is considered in this report. A clinical reappraisal study blindly re-interviewed a subsample of NCS-A respondents with the Schedule for Affective Disorders and Schizophrenia for School-Age Children Lifetime Version (K-SADS) (Kaufman *et al.*, 1997). As detailed elsewhere (Kessler *et al.*, 2009c), concordance between lifetime CIDI/SAQ and K-SADS diagnoses was good, with area under the receiver operating characteristic curve (AUC) 0.86–1.0 for mood disorders, 0.79–0.94 for anxiety disorders, 0.78–0.98 for behavior disorders, and 0.87 for any disorder.

At the end of the K-SADS interviews, clinicians completed the Children's Global Assessment Scale (CGAS; Shaffer *et al.*, 1983), a 0–100 rating of disorder-related impairment. We defined SED as having one or more 12-month DSM-IV/CIDI-SAQ mental disorders and either a CGAS $\leq$ 50, Bipolar I disorder (regardless of CGAS score), or a suicide attempt in the last 12 months (again, regardless of CGAS score). Stepwise regression analysis was carried out in the clinical reappraisal sample to predict SED from information in the CIDI. The predictors included 12-month DSM-IV/CIDI diagnoses of all Axis I disorders other than substance use disorders, summary measures of total number of disorders, information about suicidality (ideation, plans, attempts), scores on the Sheehan Disability Scale (Leon *et al.*, 1997; Sheehan, 1983), responses to the K6 scale (Kessler *et al.*, 2002), responses to questions about the number of days out of 365 in the past year when the adolescent was completely unable to carry out his or her usual daily activities because of specific disorders, information about overnight hospitalization for emotional or behavior problems in the past year, and information about intensity of outpatient treatment for emotional or behavior problems in the past year. A predicted probability of SED was calculated based on this prediction equation for each respondent in the clinical reappraisal sample. Concordance of this predicted probability with the measure of SED based on the K-SADS and CGAS in the clinical reappraisal sample was good, as indicated by an AUC of 0.85. Based on this result, the same prediction equation was used to impute a predicted probability of SED to each of the 9022 adolescents in the school sample of the larger NCS-A. This predicted probability is the outcome used in the current analysis.

The K6

The CIDI interview also included the K6 (Kessler *et al.*, 2002) a six-item scale of non-specific psychological distress that asked adolescent respondents to rate how frequently they experienced each of six symptoms of major depression and generalized anxiety disorder in the month before interview using the response options ‘never’, ‘a little of the time’, ‘some of the time’, ‘most of the time’, and ‘all of the time’. Analysis of the K6 in the NCS-A in this issue (Green *et al.*, 2010) indicates that at the individual level, the K6 is a moderate predictor of individual 12-month disorder (AUC 0.58–0.82) and is better able to predict SED associated with internalizing disorders (AUC 0.80) than with behavior disorders (AUC 0.75). Green *et al.* (2010) augmented the K6 with five additional CIDI items that specifically assessed behavior disorders. This improved the prediction of individual disorders (AUC 0.71–0.86), SED associated with internalizing disorders (AUC 0.86), and SED associated with behavior disorders (AUC 0.86). Both the original K6 and this augmented scale are used in the current study to estimate school-level SED.

Socio-demographic and school variables

The socio-demographic covariates considered here include respondent age (≤13, 14, 15, 16, 17, ≥18 years), sex, race/ethnicity (non-Hispanic White, non-Hispanic Black, Hispanic, other), and age at entrance into primary school (ages 6, 7, >7 years). School-level predictors include school size (≤50 or >50 teachers) and public/private school status.

The analysis method

We fit a two-level multilevel (hierarchical) Bayes model with bivariate outcomes to the NCS-A school sample data (Li and Zaslavsky, submitted for publication), extending previous applications of multilevel models to small-area estimation (Ghosh and Rao, 1994; Rao, 2003). The two outcomes measured for each student were the K6 (or augmented K6) score and the probit transformation of the predicted probability of SED from the model described above based on the CIDI. Predictors of each outcome included the socio-demographic and school variables listed above and random effects for school and individual respondents. The model specification (see Statistical appendix) allowed random effects for K6 and SED to be correlated at each level. We therefore estimated variance components at the school level representing the amount of variation across schools in mean K6 scores and mean SED (probit-transformed) probabilities as well as a correlation coefficient quantifying the association between the K6 and the full assessment for schools. Variance components at the individual level had similar interpretations with respect to deviations of individuals’ scores from the means at their respective schools. We fitted these models using a Bayesian procedure in MLwiN2.02 statistical software (Browne, 2005) with a conservative prior centered around 0 correlation. Further details of the model and prior distributions, the estimation procedure, model and fitting checks and diagnostics, and derivations of prediction equations are reported elsewhere (Li and Zaslavsky, submitted for publication).

The parameters of these model equations can be used to obtain school-level predictions of SED prevalence for schools in which a sample of students was administered the K6 but not the CIDI. These school-level predictions combine predictions from the regressions of SED on demographic and school characteristics with predictions based on the school-level association between the K6 and SED, weighting each according to its reliability. The reliability of a school’s mean score on the screening scale, and hence its weight in the combined prediction, depends on the size of the sample of students screened in the school.

In addition to the school-level predictions, we used estimated values of parameters to investigate the reliability of estimated rates of SED for various assumed school sample sizes representing realistic sizes of schools, extrapolating from our data. We also calculated the coefficients of the prediction equation for individual students at a hypothetical school in which screening data are available for that student as well as for a sample of the entire student body of the school. For comparison, we also applied a simpler regression-synthetic approach to small-area estimation. To calculate these predictions, we first regressed the SED score on a set of predictors, either covariates alone, covariates plus the standard K6, or covariates plus the augmented K6. We then calculated the predicted probability of SED under this fitted model for every student, and calculated the mean of those probabilities in each school as an estimate of the school’s prevalence of SED.

Results

Results of the multilevel model

Overall predicted prevalence of SED is 5.7% in the NCS-A school sample. Prediction coefficients in multilevel models for the original K6 and SED in that sample are given in Table 1. Coefficients in models with the augmented K6, although not shown, are very similar except

Table 1 Coefficients of the multilevel model for original K6 score and serious emotional disturbance (SED) score (probit transformation of predicted probability of SED)

	K6		SED	
Predictor variable	Est	(SE)	Est	(SE)
Intercept	2.643*	(0.175)	–2.382*	(0.049)
Sex[1]				
Male	–0.654*	(0.069)	–0.108*	(0.017)
Age[2]				
14 years	0.247*	(0.118)	0.082*	(0.028)
15 years	0.489*	(0.128)	0.148*	(0.032)
16 years	0.354*	(0.128)	0.147*	(0.032)
17 years	0.551*	(0.133)	0.217*	(0.033)
18 years	0.485*	(0.176)	0.196*	(0.044)
Race/Ethnicity[3]				
Black	0.607*	(0.105)	0.028	(0.027)
Hispanic	0.405*	(0.104)	0.098*	(0.027)
Other race	0.926*	(0.154)	0.090*	(0.037)
Age at starting school[4]				
Start school at age 7 years	0.242*	(0.077)	0.061*	(0.019)
Start school at age > 7 years	0.789*	(0.155)	0.190*	(0.038)
School size[5]				
School with >50 teachers	0.204*	(0.101)	0.097*	(0.029)
School type[6]				
Public school	0.300*	(0.140)	0.150*	(0.040)

*Significant at the 0.05 level, two-sided test.
[1] Comparison group = Female.
[2] Comparison group = 13-year-olds.
[3] Comparison group = Non-Hispanic White.
[4] Comparison group = Started school before age 7 years.
[5] Comparison group = School with ≤50 teachers.
[6] Comparison group = Private school.

for a difference in scale from the K6 (Li and Zaslavsky, submitted for publication). The coefficients for predicting the K6 and SED generally show similar patterns, although their magnitudes are not directly comparable because of the different scaling of the two outcome variables. Significant individual-level predictors (for both outcomes unless otherwise noted) include female sex, age greater than 13 years (with a gradient toward higher scores with increasing age for SED), Black race (for K6), Hispanic ethnicity or 'other' race, and having started school later than age 6 years. Students in larger schools and public schools also tended to have higher scores.

Random-effects variance estimates suggested that there is substantial variation among schools in both mean K6 scores and SED prevalence (Table 2) even after controlling for covariates, although individual-level variability is much greater (10.6 individual-level versus 0.3 school-level variance for the K6; 0.6 individual-level and 0.04 school-level variance for SED). To illustrate the magnitude of between-school variation, consider three schools (large enough for the effect of individual-level variability on prevalence to be negligible) with the same school characteristics and the same distributions of student covariates (age, sex, race, etc.). Suppose that one of these schools is at the median of the conditional distribution of prevalence given those characteristics (i.e. the school-level random effect is 0) and has prevalence 5.7%. Then an otherwise similar school that is one standard deviation above the median for those characteristics would have a prevalence of 7.7% and a school one standard deviation below the median would have a prevalence of 4.1%. These results suggest that considerable improvement in predictions might be expected if predictions from demographic characteristics could be supplemented with screening data that predicted the school-level random effect (i.e. the residual from the prediction based on covariates alone).

Table 2 Random-effects variance-covariance parameter estimates for mean original or augmented K6 scores and serious emotional disturbance (SED) scores

	Outcome variables	
	K6 and SED	Augmented K6 and SED
I. Individual level		
Variance (K6)	10.555	0.371
Variance (SED score)	0.597	0.597
K6-SED correlation	0.518	0.544
II. School level		
Variance (K6)	0.314	0.019
Variance (SED score)	0.036	0.037
K6-SED correlation	0.695	0.845

An important finding from the multilevel model is that the estimated correlation for concordance between the K6 and SED is very strong at the school level ($\rho = 0.70$). It is noteworthy that this association is stronger than the correlation at the individual level ($\rho = 0.52$). The augmented K6 has even stronger correlations with SED ($\rho = 0.85$ at the school level and $\rho = 0.55$ at the individual level), demonstrating the importance of additional measures of behavior disorder for predicting SED. Scatterplots of the school average SED and school average K6 and augmented K6 visually demonstrate the strength of these associations (Figure 1).

Reliability and number of respondents screened in the small area

As noted above, the reliability of school-level SED estimates depends on the number of students surveyed in the school. To illustrate this relationship, we calculated reliability using formulae derived by Li and Zaslavsky (submitted for publication). Given that the school-level association between K6 scores and SED is $\rho = 0.70$, the upper bound of reliability (i.e. reliability with a sample so large that sampling error is vanishingly small) is 0.70^2, i.e.

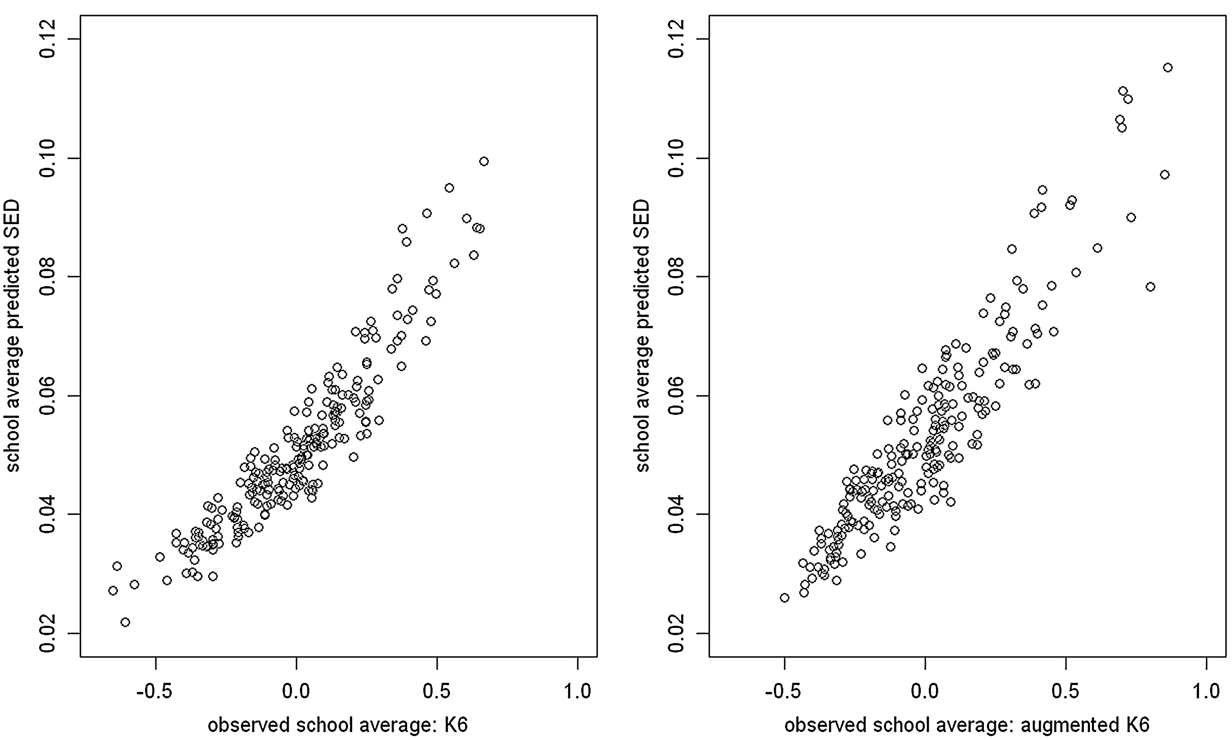

Figure 1 Scatterplot of school-level K6 and the augmented K6 versus predicted serious emotional disturbance (SED) for schools with more than 25 screened students.

0.49. The question then is how large a sample is required to approach this boundary. Our calculations show that with a sample of 25 students per school the reliability using the original K6 is 0.36, but with 200 students reliability rises to 0.44, close to the 0.49 upper limit. The upper bound of reliability with the augmented K6 is 0.72 (0.85^2), and with 200 students reliability of 0.67 is attained if they constitute the entire enrollment of the school, and 0.60 if they are a sample from a school with much larger total enrollment.

Comparison of the accuracy of multilevel model predictions to synthetic estimation

The benefits of the multilevel model predictions relative to those from synthetic estimation are illustrated by comparing prediction error in estimating school-level SED prevalence under the different modeling approaches. We made these calculations both for the actual NCS-A sample and for a simulated sample of schools with 400 students per school. The ‘relative RMSE’ columns of Table 3 compare root mean square error for each method with that from our preferred method described above, using augmented K6 data. The first three lines assess predictions under regression-synthetic models, in which we used an individual-level regression equation to predict SED and to impute school-level SED prevalence as the mean predicted probability of SED based on the individual-level prediction equation for all students sampled in the school. The conventional synthetic estimation model using individual-level SED predictions based only on socio-demographic variables has RMSE 51–60% higher than that of the preferred model, whereas the addition of the K6 to the synthetic model prediction equation reduces excess RMSE to 23–43%. Use of the augmented K6 yields even greater reduction in RMSE to only 4–8% excess. The last two rows show that use of the multilevel model further reduces error when the augmented K6 is used.

Differences among the methods are further elucidated by tabulating mean observed and predicted prevalence, grouped by quintile of predicted prevalence, as they might be in allocating resources to schools or other small areas (Table 4). As expected, observed prevalence (‘obs’ lines in the table) increases from lowest to highest quintile of predicted prevalence for each method of prediction. Importantly, the models that use the K6 show a wider spread across the quintiles for both observed and predicted prevalences (widest with augmented K6). Hence, the K6 consistently improves the discrimination between lower-prevalence and higher-prevalence schools. Comparison of predicted and observed ranges shows that the models generally under-predict prevalence in the highest quintile, much more so in the NCS-A (i.e. where the within-school samples are small) than in the simulated dataset with larger samples. Consequently the actual range across quintiles is generally understated by the model predictions. The models in their present form might be better for distinguishing among low-prevalence, average-prevalence, and high-prevalence schools than for estimating the exact prevalence in each.

Improving individual-level prediction using multilevel models

The prediction formulae used to estimate SED school-level prevalence can also be used to predict the

Table 3 Mean absolute error (MAE) and root mean squared error (RMSE) of prediction of serious emotional disturbance (SED) prevalence under five prediction models[1]

	Actual (NCS-A data)			Simulated schools with 400 students		
	MAE	RMSE	Relative RMSE[2]	MAE	RMSE	Relative RMSE[2]
Synthetic without K6	2.95	2.34	1.51	1.23	1.55	1.60
Synthetic with K6 and covariates	2.56	1.91	1.23	1.03	1.39	1.43
Synthetic with augmented K6 and covariates	2.37	1.61	1.04	0.82	1.04	1.08
Multilevel model with K6 and covariates	2.57	1.99	1.28	1.04	1.37	1.42
Multilevel model with augmented K6 and covariates	2.33	1.55	1.00	0.74	0.97	1.00

[1] All values are in units of percentage points of SED prevalence.
[2] Relative RMSE = (RMSE for this model)/(RMSE for final model), where ‘final model’ is the multilevel model with the augmented K6.

Int. J. Methods Psychiatr. Res. 19(Supplement 1): 88–98 (2010). DOI: 10.1002/mpr

Table 4 Observed and predicted mean school serious emotional disturbance (SED) prevalence in quintiles of predicted school prevalence under five prediction models[1]

		NCS-A sample sizes Quintile of predicted prevalence					Simulated with sample size 400 Quintile of predicted prevalence				
		1	2	3	4	5	1	2	3	4	5
Synthetic without K6	obs[2]	3.2	5.1	5.4	5.8	10.3	4.4	5.2	5.4	6.0	6.6
	pred[3]	4.2	4.8	5.1	5.6	6.5	4.5	5.1	5.3	5.7	6.5
Synthetic with K6 and covariates	obs[2]	2.8	4.1	5.3	6.5	11.2	3.4	4.8	5.0	6.0	7.2
	pred[3]	3.5	4.5	5.2	5.9	7.8	3.0	4.1	4.9	5.5	7.0
Synthetic with augmented K6 and covariates	obs[2]	2.5	4.0	5.0	6.6	11.7	3.3	4.6	5.2	6.1	8.5
	pred[3]	3.4	4.4	5.1	5.9	7.9	3.1	4.4	5.1	5.7	7.6
Multilevel with K6 and covariates	obs[2]	2.9	4.1	5.1	6.6	11.3	3.3	4.5	5.0	5.9	7.8
	pred[3]	3.3	4.2	4.7	5.5	7.3	3.4	4.6	5.1	5.8	7.6
Multilevel with augmented K6 and covariates	obs[2]	2.6	3.9	5.1	6.5	11.9	3.1	4.3	5.1	6.3	8.9
	pred[3]	3.1	4.0	4.8	5.5	7.9	3.2	4.5	5.2	6.0	7.9

[1] All values are in units of percentage prevalence.
[2] obs = mean of observed (for NCS-A) or simulated observed (for simulated data) prevalences for schools in a quintile of predicted prevalence under the given model.
[3] pred = mean of predicted prevalences for the corresponding collection of schools.

probability of SED for an individual student in a school where school-wide screening is conducted. This application of the prediction formula combines information from covariates, the student's own K6 score (original or augmented), and the distribution of K6 scores of other students in the school. Because the student's K6 score is only a moderately strong predictor of SED, the additional information about prevalence provided by other students' scores can have a substantial effect on the estimation of the individual's probability of SED. For example, consider two students with the same moderately high K6 score, at two large schools with respectively moderately low (1 standard deviation below average) and moderately high (1 standard deviation above) K6 means. Using the variance components estimates for the augmented K6, if the predicted probability of SED for the student at the low-prevalence school is 6.7%, the corresponding probability at the high-prevalence school would be 12.3%. Conversely, if a cut-off for follow-up screening of individual students is based on predicted probability of SED, the corresponding K6 cut-off score would be set lower at a school at which data from other students suggest a high prevalence than at a school where aggregate data suggest a low prevalence.

Discussion

The K6 has proved to be a very useful tool in screening for psychological distress at the individual level. The results of this study show that the K6 is an even more powerful predictor of SED at the school level. The brevity of the K6 means that it can be used feasibly to carry out a school-based screening survey and that the results of this survey can be used in conjunction with a Bayesian prediction system of the sort developed here to generate a useful estimate of the prevalence of SED in the school. It is reasonable to speculate that the same might be true for using the K6 to predict prevalence of SMI at the county and state levels based on data obtained in ongoing health tracking surveys. Similarly, surveys could be conducted of populations of special interest within small areas, such as minorities or the homeless.

The method presented here for small-area estimation combines two types of data – covariates (demographic and school or area characteristics) and K6 screening scale measures – and requires an equation to be estimated in a sample in which a gold-standard measure of the outcome is assessed among the same people who complete the K6. To apply these methods for geographical area estimates of SMI or for special adult populations, additional validation studies would be needed to estimate associations between K6 scores and gold-standard measures of SMI for the corresponding populations and units of analysis. The covariates can be used alone to estimate SED prevalence, as they have been in previous synthetic estimation studies of county-level and state-level SMI, but the estimates can be made much more reliable by also including K6 data.

The association of the K6 with SED was notably stronger at the school level than at the individual level. One way to understand this finding is to note that the 'signal' in the individual-level correlation of K6 with SED is diluted by the 'noise' of individual variation in responding to the screening scale. This variation tends to average out at the school level, making the school-level variation emerge more clearly as the number of students increases. Another striking result is that maximum reliability of small-area estimation can be approached with relatively small samples (200–400 respondents). It is important to note, though, that because the prediction equation for the model including the K6 variable depends on the reliability of the estimates of the mean K6, the optimal prediction coefficients will be different for schools with different-sized samples. This is because the precision, and hence the optimal relative weights, of the individual-level and aggregate-level components will change as the number of respondents surveyed in the school changes.

An important advantage of the multilevel model is that it allows us to estimate model parameters with a dataset that has relatively small sample sizes per area (i.e. the NCS-A) and then extrapolate to define a prediction equation designed to work well in areas with larger samples in which the K6 information is more reliable. Consequently, in our evaluations, the advantages of the multilevel model were more apparent in simulations of larger schools than in the NCS-A data.

We also showed, finally, that school-level prevalence data can be useful in conjunction with screening for caseness in individual students. It might seem paradoxical that probabilities of SED estimated for one student can be affected by K6 scores of other students in the same school, but the reasoning is actually quite straightforward. Consider a student who has not been screened at a school where screening has been conducted among a sample of other students. The best estimate of that student's probability of SED would be the estimated prevalence for that school, not the national mean of prevalence. Additional knowledge of the particular student's K6 score and demographic characteristics would then inform inference about whether the student is above or below the school mean. By reflecting the clustering of SED disproportionately in some schools, incorporation of school-level information into scoring of individuals tends appropriately to direct more resources (if allocated across a larger system or area) to schools with demonstrably higher prevalence than would be the case if purely individual estimation were conducted, ignoring differences among schools. On the other hand, when each school has predetermined resources for mental health services, within-school norms and cut-offs can be used to determine which students are most in need of services and prioritize these students' care.

The current study has several limitations. First, our multilevel model tends to under-predict prevalence for schools in the upper tail of the prevalence distribution, although it is fairly successful in identifying and ranking these schools. Ongoing statistical research is addressing this problem through more flexible modeling of the distribution of prevalence and simpler calibration approaches. Second, with regard to this particular application, the standard K6 is better able to detect SED associated with internalizing disorders than behavior disorders (Green *et al.*, 2010). The finding of some discordance between the K6 and SED might be partially attributable to variations in the types of SED-associated disorders prevalent in different schools. Our finding that the augmented K6 improves on the standard K6 in predicting both individual-level and school-level SED indicates that further scale refinement is needed for use with adolescents. Although the augmented K6 improves on the existing K6 for estimation of school SED prevalence, the CIDI item bank for augmenting the scale was limited to items in the NCS-A. Future K6 scale refinement for use with adolescents should focus on expanding the item list to test a wider range of symptoms of behavior disorders that can contribute to SED detection in youth. Finally, the version of the K6 considered here assessed 30-day psychological distress whereas SED was evaluated over a 12-month recall period. Given this discrepancy in timeframe, our estimates of school-level K6-SED concordance are likely conservative. A 12-month version of the K6 exists to correct this problem, assessing the same six symptoms in the worst month of the past year rather than in the past month. This worst-month version is used in the US National Household Survey on Drug Use and Health, as described in this issue by Colpe *et al.* (2010). As a result, it would be useful to apply our Bayesian multilevel regression model approach to those data to generate county-level and state-level estimates of SMI. Despite these limitations, the K6 demonstrates promise as a brief and feasible measure of small-area SED or SMI that can provide comparative information about need for resource allocation while also improving estimation for individuals.

Statistical appendix

We assume a bivariate generalized linear mixed model for the NCS-A data:

$$Y_{ijm} = X_{ijm}\beta_m + v_{im} + e_{ijm}, \text{ for } m = 1, 2,$$

where m stands for the type of outcome with $m = 1$ referring to K6 or augmented K6 and $m = 2$ referring to SED; Y_{ijm}, X_{ijm} stand for the outcome of type m and corresponding measured covariates for subject j in school i; v_{im} is the school-specific random effect for outcome m in school i; e_{ijm} is the random error for outcome m of individual j in school i. Moreover, we assume normal random effects distributions $(v_{i1}, v_{i2})' \sim N(0, \Sigma_v)$, $(e_{ij1}, e_{ij2})' \sim N(0, \Sigma_e)$, with covariance matrices:

$$\Sigma_v = \begin{pmatrix} \sigma_{v1}^2 & \rho_v\sigma_{v1}\sigma_{v2} \\ \rho_v\sigma_{v1}\sigma_{v2} & \sigma_{v2}^2 \end{pmatrix}, \quad \Sigma_e = \begin{pmatrix} \sigma_{e1}^2 & \rho_e\sigma_{e1}\sigma_{e2} \\ \rho_e\sigma_{e1}\sigma_{e2} & \sigma_{e2}^2 \end{pmatrix},$$

where ρ_v, ρ_e are the correlations between random effects of the two outcomes respectively at the school and individual levels.

The above model is estimated through a hierarchical Bayes approach, with non-informative or weakly informative prior distributions for the parameters $(\beta, \Sigma_v, \Sigma_e)$, and parameter estimates are reported as posterior means and credible intervals.

Acknowledgements

The National Comorbidity Survey Replication Adolescent Supplement (NCS-A) is supported by the National Institute of Mental Health (NIMH; R01-MH66627 [Zaslavsky] and U01-MH60220 [Kessler]) with supplemental support from the National Institute on Drug Abuse (NIDA), the Substance Abuse and Mental Health Services Administration (SAMHSA), the Robert Wood Johnson Foundation (RWJF; Grant 044780), and the John W. Alden Trust. The authors thank Michael J. Gruber and Nancy A. Sampson for their help in preparing the dataset for analysis. The views and opinions expressed in this report are those of the authors and should not be construed to represent the views of any of the sponsoring organizations, agencies, or U.S. Government. A complete list of NCS-A publications can be found at http://www.hcp.med.harvard.edu/ncs. Send correspondence to ncs@hcp.med.harvard.edu. The NCS-A is carried out in conjunction with the World Health Organization World Mental Health (WMH) Survey Initiative. We thank the staff of the WMH Data Collection and Data Analysis Coordination Centres for assistance with instrumentation, fieldwork, and consultation on data analysis. The WMH Data Coordination Centres have received support from NIMH (R01-MH070884, R13-MH066849, R01-MH069864, R01-MH077883), NIDA (R01-DA016558), the Fogarty International Center of the National Institutes of Health (FIRCA R03-TW006481), the John D. and Catherine T. MacArthur Foundation, the Pfizer Foundation, and the Pan American Health Organization. The WMH Data Coordination Centres have also received unrestricted educational grants from Astra Zeneca, Bristol-MyersSquibb, Eli Lilly and Company, GlaxoSmithKline, Ortho-McNeil, Pfizer, Sanofi-Aventis, and Wyeth. A complete list of WMH publications can be found at http://www.hcp.med.harvard.edu/wmh/.

Declaration of interests

Dr. Kessler has been a consultant for GlaxoSmithKline Inc., Kaiser Permanente, Pfizer Inc., Sanofi-Aventis, Shire Pharmaceuticals, and Wyeth-Ayerst; has served on advisory boards for Eli Lilly & Company and Wyeth-Ayerst; and has had research support for his epidemiological studies from Bristol-Myers Squibb, Eli Lilly & Company, GlaxoSmithKline, Johnson & Johnson Pharmaceuticals, Ortho-McNeil Pharmaceuticals Inc., Pfizer Inc., and Sanofi-Aventis. The remaining authors report no conflicts of interest.

References

Browne W.J. (2005). *MCMC Estimation in MLwiN (Version 2.0)*, University of Bristol, Centre for Multilevel Modeling.

Burns B.J., Costello E.J., Angold A., Tweed D., Stangl D., Farmer E.M., Erkanli A. (1995). Children's mental health service use across service sectors. *Health Affairs (Millwood)*, **14**, 147–159, DOI: 10.1377/hlthaff.14.3.147.58

Colpe L.J., Barker P.R., Karg R.S., Batts K.R., Morton K.B., Gfroerer J.C., Stolzenberg S.J., Cunningham D.B., First M.B., Aldworth J. (2010). The National Survey on Drug Use and Health Mental Health Surveillance Study: calibration study design and field procedures. *Int J Methods Psychiatr Res*, **19**(Suppl. 1), 36–48.

Ghosh M., Rao J.N.K. (1994). Small area estimation: an appraisal. *Stat Sci*, **9**, 55–93, DOI: 10.1214/ss/1177010647

Goldsmith H.F., Holzer C.E., Manderscheid R.W. (1998). Neighborhood characteristics and mental illness. *Eval Program Plann*, **21**, 211–225, DOI: 210.1016/S0149-7189(1098)00012-00013

Green J.G., Gruber M.J., Sampson N.A., Zaslavsky A.M., Kessler R.C. (2010). Improving the K6 short scale to predict serious emotional disturbance in adolescents in the USA. *Int J Methods Psychiatr Res*, **19**(Suppl. 1), 23–35.

Hudson C. (2009). Validation of a model for estimating state and local prevalence of serious mental illness. *Int J Methods Psychiatr Res*, **18**, 251–264.

Kaufman J., Birmaher B., Brent D., Rao U., Flynn C., Moreci P., Williamson D., Ryan N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry*, **36**, 980–988, DOI: 10.1097/00004583-199707000-00021

Kessler R.C., Andrews G., Colpe L.J., Hiripi E., Mroczek D.K., Normand S.L., Walters E.E., Zaslavsky A.M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med*, **32**, 959–976, DOI: 10.1017/S0033291702006074

Kessler R.C., Avenevoli S., Costello E.J., Green J.G., Gruber M.J., Heeringa S., Merikangas K.R., Pennell B.E., Sampson N.A., Zaslavsky A.M. (2009a). Design and field procedures in the US National Comorbidity Survey Replication Adolescent Supplement (NCS-A). *Int J Methods Psychiatr Res*, **18**, 69–83, DOI: 10.1002/mpr.279

Kessler R.C., Avenevoli S., Costello E.J., Green J.G., Gruber M.J., Heeringa S., Merikangas K.R., Pennell B.E., Sampson N.A., Zaslavsky A.M. (2009b). National Comorbidity Survey Replication Adolescent Supplement (NCS-A): II. Overview and design. *J Am Acad Child Adolesc Psychiatry*, **48**, 380–385, DOI: 10.1097/CHI.0b013e3181999705

Kessler R.C., Avenevoli S., Green J., Gruber M.J., Guyer M., He Y., Jin R., Kaufman J., Sampson N.A., Zaslavsky A.M. (2009c). National Comorbidity Survey Replication Adolescent Supplement (NCS-A): III. Concordance of DSM-IV/CIDI diagnoses with clinical reassessments. *J Am Acad Child Adolesc Psychiatry*, **48**, 386–399, DOI: 10.1097/CHI.0b013e31819a1cbc

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. *Arch Gen Psychiatry*, **60**, 184–189.

Kessler R.C., Berglund P.A., Walters E.E., Leaf P.J., Kouzis A.C., Bruce M.L., Friedman R.M., Grosser R.C., Kennedy C., Kuehnel T.G., Laska E.M., Mandersheid R.W., Narrow W.E., Rosenheck R.A., Schneier M. (1999). Population-based analyses: a methodology for estimating the 12-month prevalence of serious mental illness. In: *Mental Health, United States 1998* (eds Manderscheid RW, Henderson MJ), pp. 99–109, U.S. Government Printing Office.

Kessler R.C., Green J.G., Gruber M.J., Sampson N.A., Bromet E., Cuitan M., Furukawa T.A., Gureje O., Hinkov H., Hu C., Lara C., Lee S., Mneimneh Z., Myer L., Oakley-Brown M., Posada-Villa J., Sagar R., Viana M.C., Zaslavsky A.M. (2010). Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) Survey Initiative. *Int J Methods Psychiatr Res*, **19**(Suppl. 1), 4–22.

Kessler R.C., Merikangas K.R. (2004). The National Comorbidity Survey Replication (NCS-R): background and aims. *Int J Methods Psychiatr Res*, **13**, 60–68, DOI: 10.1002/mpr.166

Kessler R.C., Üstün T.B. (2004). The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int J Methods Psychiatr Res*, **13**, 93–121, DOI: 10.1002/mpr.168

Leaf P.J., Alegria M., Cohen P., Goodman S.H., Horwitz S.M., Hoven C.W., Narrow W.E., Vaden-Kiernan M., Regier D.A. (1996). Mental health service use in the community and schools: results from the four-community MECA Study. Methods for the Epidemiology of Child and Adolescent Mental Disorders Study. *J Am Acad Child Adolesc Psychiatry*, **35**, 889–897, DOI: 10.1097/00004583-199607000-00014

Leon A.C., Olfson, M., Portera, L., Sheehan, D.V. (1997). Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *Int J Psychiatry Med*, **27**, 93–106, DOI: 10.2190/T8EM-C8YH-373N-1UWD

Rao J.N.K. (2003). *Small Area Estimation*, John Wiley and Sons.

Shaffer D., Gould M.S., Brasic J., Ambrosini P., Fisher P., Bird H., Aluwahlia S. (1983). A children's global assessment scale (CGAS). *Arch Gen Psychiatry*, **40**, 1228–1231.

Sheehan D.V. (1983). *The Anxiety Disease*, Scribner.

Slade E.P. (2003). The relationship between school characteristics and the availability of mental health and related health services in middle and high schools in the United States. *J Behav Health Serv Res*, **30**, 382–392.

Substance Abuse and Mental Health Services Administration. (1993). Final notice establishing definitions for (1) children with a serious emotional disturbance, and (2) adults with a serious mental illness. *Fed Register*, **58**, 29422–29425.

Thomas K.C., Ellis A.R., Konrad T.R., Holzer C.E., Morrissey J.P. (2009). County-level estimates of mental health professional shortage in the United States. *Psychiatr Serv*, **60**, 1323–1328, DOI: 10.1176/appi.ps.60.10.1323