

Estimating the Prevalence of Childhood Serious Emotional Disturbance (SED) within a National Survey: Pilot Study Experiences

Heather Ringeisen, Director, Center for Behavioral Health and Development,
Survey Research Division, RTI International

Jeremy Aldworth, Senior Research Statistician, Social, Statistical &
Environmental Sciences, RTI International

The SED Pilot Study: A Summary

Heather Ringeisen, Director, Center for Behavioral Health and
Development, Survey Research Division, RTI International

Background and Definition of Serious Emotional Disturbance (SED)

- In 1992, the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA) Reorganization Act established a block grant for Community Mental Health Services to be administered by SAMHSA's Center for Mental Health Services.
- This block grant allows funds to be allocated to States to support the provision of services to children with serious emotional disturbance (SED) and adults with serious mental illness (SMI).
- SAMHSA convened a technical advisory group that developed a definition of SED, published on May 20, 1993. This *Federal Register* definition is to be used to identify and estimate the size of the SED population within each State.

SED Definition in 1993 *Federal Register*

- Children from birth up to age 18
- Who currently or any time during the past year have had a diagnosable mental, behavioral, or emotional disorder of sufficient duration to meet diagnostic criteria specified within [the then current] DSM III R
- Which has resulted in functional impairment which substantially interferes with or limits the child's role or functioning in family, school, or community activities

Purpose of SED Definition

- Define children eligible to receive services funded by Federal community mental health services block grant dollars.
- Estimate the level of service needs by individual States to determine block grant funding levels.
- Distinguish SED from conditions covered by other block grant funding streams, such as substance use disorders and developmental disorders.

SAMHSA Requirements

- SAMHSA currently provides State-level estimates of SED for use in block grants; however, there is no recent large-scale epidemiological study that collects these data.
- To estimate the prevalence of SED within a large-scale epidemiological survey, there is need for
 - An operationalized definition of the term "SED" for use in generating national and State estimates and
 - A standard method of computing SED prevalence estimates.

Two Possible Ways to Estimate SED

1. Design-based method conducting a new and separate study
 - This option has issues determining the optimal frequency of estimation (e.g., annual vs. periodic vs. one-time).
2. Model-based method for use in any existing national survey
 - This option would still need to be developed and consequently might need original/new data collection.

Preliminary SAMHSA efforts have focused on a creating a model-based structure.

Developing a Model-Based Method to Estimate the Prevalence of SED in a National Study

- In 2006, SAMHSA convened an advisory panel that recommended the following:
 - The National Health Interview Survey (NHIS) continue to collect the five-item version of the Strengths and Difficulties Questionnaire (SDQ) for use as a predictor variable.
 - However, the expert panel noted the lack of a reliable cutoff score for this version of the SDQ (used only in the NHIS) and proposed developing a predictive model based on the 5-item SDQ and a standard psychiatric measure.
- As a result, from 2009 to 2012, SAMHSA and later the National Institute of Mental Health (NIMH) funded a pilot study to evaluate the estimation of SED in collaboration with the National Center for Health Statistics (NCHS).

Purpose of SED Pilot Study

- Primary Objective: To explore the development of a model-based procedure to estimate SED using data collected on the NHIS.
- The methodology was similar to that used to estimate Serious Mental Illness (SMI) using data collected on the NSDUH.
- The SED method was based on developing predictive models where the five-item SDQ and a sixth "impact" item from the SDQ contained in the NHIS questionnaire were the independent variables that predicted SED status (determined based upon the 'gold standard' clinical interview data).
- The pilot study was conducted under contract by RTI International in partnership with Duke University.

Pilot Study Design Summary: Sample

- Sample: Parents of children aged 4 to 17 years (who had participated in the final 3 quarters of the 2011 or first quarter of the 2012 NHIS) and youth 12 to 17 years old
 - Only sampled children whose parents completed the NHIS interview in English, provided complete contact information and SDQ responses, and indicated the child had no history of intellectual disability, developmental delay, autism, or Down syndrome were eligible for participation.
 - Sampling strata were defined by NHIS SDQ scores and then sampled proportionally to the size of the standard error of a proxy measure of child mental disorder distributed across the strata.
- A total of 217 parents (with children 4 to 17 years old) completed the pilot.
 - 139 completed parent interviews for children aged 4–11 and 78 completed parent/child pairs for children aged 12–17. Another 50 parents of 12–17 year old children responded, but the children did not.
 - Of the 1,187 identified parent respondents, 195 were ineligible due to a competing study using the NHIS sample, 277 were not locatable and 239 not contactable; and 200 refused or broke off the interview. The final sample size was 217 (18.3%).

Pilot Study Design Summary: “Gold Standard” Clinical Interview

- Interviewers administered the shortened Child and Adolescent Psychiatric Assessment (CAPA; Angold & Costello, 2000) for children 9 to 17 years, or a shortened version of the Preschool-Age Psychiatric Assessment (PAPA; Egger et al., 2006) for children 4 to 8 years by telephone to the parents or guardians of sampled children.
- Youths aged 12 to 17 were administered the shortened child CAPA interview by telephone.
- These shortened versions had only five modules: depression, anxiety, attention deficit-hyperactivity disorder (ADHD), oppositional defiant disorder, and conduct disorder.
- Consistent with prior research, the SDQ pilot study used parent report only for cases 4 to 11 years old; parent and adolescent report were used for cases 12 to 17 years old.
- The CAPA and PAPA instruments include an incapacities module that assesses impairment for all endorsed symptoms. The interviewer is asked to distinguish three levels of impaired functioning—absent, partial or severe.
- SED status (positive or negative) was determined based upon responses to the five CAPA/PAPA modules as well as scores on the incapacities module

Pilot Study Design Summary: Analyses

- A series of preliminary modeling analyses were conducted to determine issues associated with modeling SED status using independent variables from the NHIS national survey.
- The analyses modeled three different possible definitions of SED based upon varying impairment cut-points.

Variable	Definition
SED Definition 1	Presence of any disorder assessed <u>and</u> any partial or severe impairment rating
SED Definition 2	Presence of any disorder assessed <u>and</u> at least three partial or one severe impairment rating
SED Definition 3	Presence of any disorder assessed <u>and</u> any severe impairment rating

Pilot Study Design Summary: Analyses (continued)

- Models examined various ways to use the NHIS SDQ to predict SED
 - Total 5-item SDQ score
 - Five individual SDQ item scores (used with and without the 6th “impact” SDQ item. The 6th item was used as a proxy indicator of “impairment.”)
- Pilot analyses also examined the impact of age on the calibration models; however, due to small sample sizes, the results of these exploratory analyses were ambiguous.

Pilot Study Challenges

- IRB/ERB Approval Delays
 - The pilot study was delayed more than 18 months due to lengthy negotiations about necessary ethical and security requirements.
 - Key issues included the following: (1) the appropriateness of interviewing children about mental health issues by telephone, (2) pragmatic operational issues about whether (and how) to report individual results to parent and/or child participants, and (3) the determination of an appropriate protocol for handling distressed child respondents during this type of telephone-based study.
 - All of these issues lacked a clear precedent protocol to explicitly guide ERB (NCHS), IRB (RTI), and study leadership recommendations.

Pilot Study Challenges (Continued)

- Lag time between original NHIS interview and follow-up study implementation
 - Due to delays, the time lag between the original NHIS interview and the pilot study clinical interview was as long as 10-11 months for some cases.
 - Because the SDQ has a 6-month reference period and because the child's emotional state could easily have changed during this time lag, the SDQ was re-administered to parents at the time of the clinical interview.
 - Analyses indicated that SDQ responses between the time points of the original NHIS interview and the subsequent clinical interview were significantly correlated and somewhat in agreement; however, the correlation and agreement statistics were also significantly different from 1.
 - SDQ responses from the clinical interview were more predictive of the CAPA/PAPA-based SED than those from the original NHIS interview.

Pilot Study Challenges (Continued)

- Response rates
 - The response rates in this study were substantially lower than expected.
 - The time delay was one key factor. The rate of unlocatable cases (22.6 percent) where a respondent's telephone number could not be confirmed was especially high.
 - There was also a particularly high rate of ineligible/incapable cases (19.9 percent) due to the inadvertent delivery from the NCHS to RTI of cases targeted for participation in the Medical Expenditure Panel Survey (MEPS).
 - Because of the small sample size and potential bias due to the low response rate of this pilot study, we have recommended caution in extrapolating the pilot study results beyond this context.

Pilot Study Challenges (Continued)

- Reconciling different age-dependent measures and definitions of SED
 - For the 4 to 7 and 8 to 11 age groups, a positive diagnosis of any disorder was based on the parent interview alone
 - For the 12 to 17 age group a positive diagnosis was based on any of the following situations: (1) a positive diagnosis based on the parent interview, (2) a positive diagnosis based on the child interview, or (3) a positive diagnosis based on a joint determination (using the CAPA scoring algorithm, it was possible for the joint determination to be positive while both parent and child determinations were separately negative).

Age Group	Clinical Interview Measure	Parent Interview Required?	Child Interview Required?
4 to 7	Preschool Age Psychiatric Assessment (PAPA)	Yes	No
8 to 11	Child and Adolescent Psychiatric Assessment (CAPA)	Yes	No
12 to 17	Child and Adolescent Psychiatric Assessment (CAPA)	Yes	Yes

Issues Continuing to Need Resolution Prior to Generating Model-Based SED Estimates

- **Matching reference periods between screening and diagnostic assessment.** The SDQ asks about behaviors exhibited in the past 6 months; the CAPA/PAPA asks about symptoms in the past 3 months. This created a mismatch between the screening and clinical interview reference periods. Furthermore, the Federal Register definition of SED requires the presence of a “past year” mental disorder.
- **Impairment definition consensus and established cut-points.** There is no consensus for how "serious impairment" should be measured or defined as an explicit cut-point within particular measures. In the pilot study, three definitions of SED were examined, each varying in the degree of severity required to meet the definition of "serious impairment." These definitions each resulted in very different SED estimates across age groups.

Issues Continuing to Need Resolution Prior to Generating Model-Based SED Estimates

- Instrument limitations
 - **Age representation.** The 1993 *Federal Register* definition of SED is designed to encompass children birth to 18 years; however, complementary diagnostic and impairment tools do not exist to measure the presence of an impairing mental disorder across this entire age span.
 - **DSM-5 updates.** Most clinical interview tools have not yet released updates for DSM-5; some instruments will not have a DSM-5 update available.
 - **Administration mode testing beyond field data collection.** There are no studies that directly compare telephone to in-person administration for the leading lay-administered child diagnostic interviews (e.g., CIDI, DISC or CAPA).

Lessons Learned from SED Pilot Study

- The NSDUH modeling experience for adult Serious Mental Illness (discussed shortly) included a single measure of Serious Mental Illness for all adult age groups.
- *Meanwhile...*the child SED pilot study required 3 distinct SED measures that varied by child age group. This required 3 separate statistical models.
- Some lessons learned for consideration in future studies include:
 - A sufficient sample size will be needed for each age group to support age-specific models
 - Additional thought will need to be given to reconciling discontinuities between the 3 age groups (and their implications for estimating SED across ages)
 - The ramifications of different types/numbers of reporters for statistical modeling needs to be better understood. In the pilot study, the oldest age group required both parent and child responses for the clinical interview (consistent with field standards), but only parent responses were available in NHIS questionnaire.

Lessons Learned from CBHSQ model-based estimation efforts of SED and SMI

Jeremy Aldworth, Senior Research Statistician, Social, Statistical &
Environmental Sciences, RTI International

Model-Based Method of Estimating SED: Background

- Gold Standard measure of SED = diagnosis from clinical interview (Yes/No)
- In practice – Gold Standard measure too expensive/time consuming to obtain from all respondents in a large survey
- Therefore – need to apply (cheaper, more convenient) model-based method to obtain SED estimates
- For model-based method – condition on Gold Standard measure = Truth
 - Difference between Gold Standard measure and “real” Truth not accounted for
 - Model-based bias and error measures use Gold Standard measure as the reference (truth)

Model-Based Method of Estimating SED: Nutshell

- Model-based method in a nutshell:
 - Administer SED-predicting scale(s) (e.g., SDQ) to all eligible respondents in main survey
 - Select subsample from eligible respondents in main survey and administer clinical interview to obtain gold standard measure of SED among selected respondents
 - Use logistic regression model to “match” gold standard measure of SED to scale(s) administered in questionnaire within subsample (because both measures available)
 - Gold standard measure = response variable
 - SED-predicting scale(s) = predictor variable(s)
 - May include other covariates (associated with SED) in model
 - Select cutpoint to dichotomize predicted probabilities of SED based on model into Yes/No SED predictions
 - Extrapolate model (and cutpoint) to all eligible respondents in main survey to determine SED status (Yes/No) for each respondent, and then can provide prevalence estimates (overall and within demographic and other subgroups)

Model-Based Method of Estimating SED: Details

- In subsample – Gold standard measure and scale responses available
- Let

$$Y = \begin{cases} 0 & \text{if gold std SED negative} \\ 1 & \text{if gold std SED positive} \end{cases}$$

$$\pi = \Pr(Y = 1 \mid \mathbf{x}); \quad \mathbf{x} = \text{Scale data}$$

$$\text{logit}(\pi) = \mathbf{x}'\boldsymbol{\beta}; \quad \boldsymbol{\beta} = \text{Population regression coefficients}$$

- Then each respondent has predicted probability of SED:

$$\hat{\pi} = \text{logit}^{-1}(\mathbf{x}'\hat{\boldsymbol{\beta}}) \quad (\text{estimated with sample weights})$$

Model-Based Method of Estimating SED: Details (2)

- In subsample, we have

$$Y = \begin{cases} 0 & \text{if gold std SED negative} \\ 1 & \text{if gold std SED positive} \end{cases}$$

$$\hat{\pi} = \text{logit}^{-1}(\mathbf{x}'\hat{\boldsymbol{\beta}}) \in (0,1) \text{ interval}$$

- We need predicted probability of SED to be dichotomized:

$$\hat{Y} = \begin{cases} 0 & \text{if } \hat{\pi} < \pi_0 & \text{= Predicted SED negative} \\ 1 & \text{if } \hat{\pi} \geq \pi_0 & \text{= Predicted SED positive} \end{cases}$$

π_0 = *Cutpoint with no or minimum possible (absolute) bias*

Model-Based Method of Estimating SED: Details (3)

- Receiver operating characteristic (ROC) analysis used to determine minimum-biased cutpoint (for each candidate model)
- Minimum-biased cutpoint (π_0) determined by coming as close as possible to equalizing (offsetting) false positive ($Y = 0, \hat{Y} = 1$) and false negative ($Y = 1, \hat{Y} = 0$) counts
- Some useful statistics from ROC analysis:
 - False positive (negative) rate = False positive (negative) count/total count
 - Model-based bias (rate) = false positive rate – false negative rate
 - Classification error (rate) = false positive rate + false negative rate
- Cutpoint that minimizes bias for all youths may not minimize bias within subgroups
- Ideal Goal: Select model that minimizes classification error subject to minimum bias (overall and within demographic and other subgroups)

Model-Based Method of Estimating SED: Purpose

- Purpose of model-based method – to provide prevalence estimates overall and within subgroups, and not as an individual diagnostic test
- Therefore – minimization of bias crucial (error rate not as important)
- For individual diagnostic test – needs are very different and depend on disease
- Ex: diagnostic test for Ebola:
 - False negative = death sentence
 - False positive = cost, inconvenience
 - Therefore – critical for false negative rate to be near 0, irrespective of false positive rate (i.e., bias not important)

Lessons Learned from MHSS (SMI): Impact of Weights

- MHSS = Mental Health Surveillance Study to estimate SMI (“adult SED”)
- National estimates of SMI (SED) – hence also model-based analyses – based on weighted data
- Optimal design to select subsample (e.g., Neyman allocation) may result in large weights relative to other weights

Lessons Learned from MHSS (SMI): Weights (2)

- Large weights may have undue influence in determining position of cutpoint – maybe OK for subsample, but maybe not OK for extrapolation to main survey (which is used to provide prevalence estimates)
- Large weight “straddling” neighborhood of cutpoint could make it difficult to come close to equalizing false positive/false negative counts – resulting in bias
- “Optimal” sample design for selecting subsample could generate great variability in weights – may need to modify this!
- In MHSS – subsequently modified subsample design by “reversing” selected over/undersampling in main survey – reduced weight variability

Lessons Learned from MHSS (SMI): Sample Size

- Sample size of subsample impacts gold standard estimates, and model-based analyses and estimates
- Impact on gold standard estimates:
 - Gold standard measure is assumed to be “true” for each selected respondent (conditional)
 - But – If sample size of subsample too small, gold standard prevalence estimates may be subject to large design-based sampling error (particularly at the subgroup level) – i.e., subsample prevalence estimates may differ considerably from those obtained from a census (i.e., population prevalence estimates – with no sampling error)
 - Even “perfect” model cannot address design-based sampling error – model is used to “match” gold standard prevalence estimates made available by sample design

Lessons Learned from MHSS (SMI): Sample Size (2)

- Impact on model-based analyses and estimates if subsample size too small:
 - Weighted model-based analyses and estimates subject to same design-based sampling error (again, particularly at the subgroup level)
 - Model-based bias and classification error conditional on subsample prevalence estimates
 - Model-based errors:
 - May not have enough data to identify “best” model (model misspecification error – could result in unmeasured bias at subgroup level)
 - Beta estimates of model may have large SEs (model error)
- Ex: MHSS:
 - After 1 year, n = 750 allowed determination of 2 DF model (with K6, WHODAS scales)
 - After 5 years, n = 5,000 allowed determination of improved 5 DF model (with K6, WHODAS scales, Age variable, PY MDE, and PY Suicidal Thoughts)
 - Age variable – large reduction in bias within Age Groups
 - PY MDE, PY Suicidal Thoughts – large reduction in classification error
 - Example of better specified model, and model error was reduced (betas had smaller SEs)

Lessons Learned from MHSS (SMI): Sample Size (3)

- Model misspecification error and model error will be transferred to error in prevalence estimates
- BUT not accounted for in SEs associated with model-based prevalence estimates obtained from the main-study data
 - Technical difficulties in accounting for model-based errors in SEs of prevalence estimates – so added note that prevalence estimates and SEs are conditional on model
 - Less of a problem if sample size of subsample sufficiently large

Lessons Learned from MHSS (SMI): Adding Covariates

- “Improved” SMI model included PY MDE, PY Suicidal Thoughts – substantially decreased classification error
- However:
 - Problems arose in joint analyses of SMI and PY MDE or PY Suicidal Thoughts
 - Ex: estimate of proportion of adults with PY MDE who also had SMI
 - Investigation indicated prevalence estimates (as in Ex above) tend to be overestimated
- When formulating final model – need to take into account list of variables assumed to be important in terms of planned joint analyses with SMI/SED
 - If any variables on list are important covariates in model, may consider developing alternative versions of final model that exclude such covariates specifically for joint analyses involving those variables

Lessons Learned from MHSS (SMI): Other Levels of MI

- Primary MI variable = Serious MI (SMI)
- Secondary MI variables = Moderate and Mild MI (MMI and LMI)
- How to model all levels of MI (SMI/MMI/LMI) while maintaining primacy of SMI?
 - Polytomous model (4-level MI variable) – Optimized over all levels, not SMI; complicated
 - Separate models – Simple, but consistency not guaranteed
 - SMI model with different cutpoints for MMI/LMI – Simple, consistent, primacy of SMI maintained, appeared to provide reasonable estimates of MMI/LMI (although not optimized for these levels)

Lessons Learned from MHSS (SMI): Others

- Logistic regression fit may not be best, particularly in the tails – BUT may not be a problem since dichotomization by cutpoint merely requires model to provide best ordering of predicted probabilities
 - Ideal ordering : $Y=1$ cases to have large predicted probabilities, $Y=0$ cases to have small predicted probabilities
- Predicted probabilities are (theoretically) continuous in $(0,1)$ interval – BUT in practice have only as many distinct values as there are in all combinations of the predictor variables
 - Ex: If only 1 predictor variable with 6 levels, then only 6 distinct predicted probabilities possible
 - If too few distinct predicted probabilities, gradations between them may be too coarse to identify minimum-biased cutpoint (e.g., bias can merely switch signs between gradations)
 - In practice – cutpoint can occupy any value within interval between gradations (consecutive predicted probabilities) – some indeterminacy if interval large

Lessons Learned from MHSS (SMI): Level of Effort!

- Substantial level of effort (LOE) in terms of work, time, intensity went into making decisions about MHSS and then in its implementation to estimate SMI
- Similar LOE required for study to estimate SED – if not greater, due to extra complications among children/youths relative to adults (e.g., multiple measures by age group among children/youths vs. single SMI measure)

Summary: Pending Next Steps for SED Prevalence Estimation

- Develop a well-operationalized definition of SED, amenable to estimation in a national study (including recommended cut-points for various age groups on available impairment measures).
- Operationalize existing measures of impairment with concrete, developmentally grounded and culturally sensitive anchors to increase the accuracy of their assessment. Develop tools to assess impairment in very young children.
- Build agreement around best predictive and gold standard measures to use in estimating the national and state prevalence of SED from birth to 18 years.
- Leverage existing datasets that might support analyses to assess the power of various candidate tools to predict SED in statistical models.

Summary: Pending Next Steps for SED Prevalence Estimation (continued)

- Provide suggestions for how to address study seam effects that result from changes in an instrument type or the required number of reporters by child age.
- Provide suggestions about study data collection mode (telephone, in-person, etc) that consider both data quality and cost efficiency.
- Provide suggestions for how to address the varying reference periods that exist between the Federal Register definition of SED, predictive tools embedded in candidate national surveys, and “gold standard” clinical assessment tools.