



# HHS Public Access

Author manuscript

*Int J Methods Psychiatr Res.* Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

*Int J Methods Psychiatr Res.* 2015 December ; 24(4): 266–274. doi:10.1002/mpr.1471.

## Estimating the Prevalence of Any Impairing Childhood Mental Disorder in the National Health Interview Survey

**Heather Ringeisen, Ph.D.,**

RTI International, Research Triangle Park, NC

**Jeremy Aldworth, Ph.D.,**

RTI International, Research Triangle Park, NC

**Lisa J. Colpe, Ph.D., M.P.H.,**

National Institute of Mental Health, Bethesda, MD

**Beverly Pringle, Ph.D., and**

National Institute of Mental Health, Bethesda, MD

**Catherine Simile, Ph.D.**

National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD

### Abstract

This study investigates whether the 6-item Strengths and Difficulties Questionnaire SDQ (5 symptoms and one impact item) included in the National Health Interview Survey can be used to construct models that accurately estimate the prevalence of any impairing mental disorder among children 4–17 years old as measured by a shortened Child/Adolescent or Preschool Age Psychiatric Assessment (CAPA or PAPA). A subsample of 217 NHIS respondents completed a follow-up CAPA or PAPA interview. Logistic regression models were developed to model presence of any child mental disorder with impairment (MDI) or with severe impairment (MDSI). Models containing only the SDQ impact item exhibited highly biased prevalence estimates. The best-performing model included information from both the five symptom SDQ items and the impact item, where absolute bias was reduced and sensitivity and concordance were increased. This study illustrates the importance of using all available information from the 6-item SDQ to accurately estimate the prevalence of any impairing childhood mental disorder from the NHIS.

### Keywords

children; adolescents; epidemiology; estimation; methods

---

Correspondence: Heather Ringeisen, Ph.D., Survey Research Division, RTI International, P.O. Box 12194, Research Triangle Park, NC 27709. hringeisen@rti.org Voice: (919) 541-6931 Fax: (919) 541-1261.

**Competing interests:** The authors have no competing interests.

## Introduction

In 1998, key recommendations in a report by a National Advisory Mental Health Council (NAMHC) board called for an ongoing national survey to monitor youth's mental health disorders, impairment, and service use (U.S. Department of Health and Human Services, 2000). More than a decade later, this goal is only partially realized. Assessment of mental disorders during childhood and adolescence is challenging. Typical behavior during these developmental stages ranges widely. Early childhood assessments can be the most challenging, partly explaining why there are comparatively few national estimates of mental disorders among children younger than 8 years old. While two national studies provide prevalence estimates of specific mental disorders for children 8–15 years (Merikangas et al., 2010) and 13–17 years (Kessler et al., 2012) neither provides ongoing surveillance of mental disorders. Certain national surveys, such as the National Health Interview Survey (NHIS), produce estimates of select disorders based on parent report (e.g., ADHD); however, these surveys do not include estimates of any childhood mental disorder using a standardized tool.

In response to the NAMHC recommendation, the National Institutes of Health National Institute of Mental Health (NIMH) partnered with the Centers for Disease Control and Prevention's National Center for Health Statistics (NCHS) to incorporate variations of the Strengths and Difficulties Questionnaire (SDQ), a widely used tool to screen for child mental health problems, into the NHIS in 2001 (Bourdon et al., 2005). The SDQ extended version (25 symptom items; 5 impact items) was included in the 2001, 2003, and 2004 NHIS; the 6-item version of the SDQ (5 symptom items; 1 impact item) was included in 2002, 2005–2007 and 2010–2013; and the impact item alone appeared in the 2008–2009 NHIS. NHIS data are cited regularly in federal reports such as *America's Children: Key National Indicators of Well-Being* (Federal Interagency Forum on Child and Family Statistics, 2012). SDQ data have been analyzed in multiple ways over the past decade. Bourdon et al. (2005) established U.S. normative scoring bands for the extended SDQ, defining children with high difficulties as those in the top 10% band. The researchers then used a measure of mental health service contact to test the predictive validity of SDQ scores and found that, indeed, higher SDQ scores correlated with greater service use. Pastor and colleagues (2012) examined (6-item) versions of the SDQ and used “high symptom score” (i.e., total SDQ score  $\geq 6$  from the 5 symptom items) and/or “serious overall difficulties” (i.e., report of definite or severe difficulties on the impact item) as indicators of mental health problems. They identified a total of 7.4% of children as having mental health problems, using a combination of scoring methods: 2.1% had both high symptom and serious overall difficulties scores, 2.2% had a high symptom score alone, and 3.1% had only the serious overall difficulties rating. Both Bourdon and Pastor used service use and other child characteristics to compare children who were rated as having or not having mental health problems – a criteria-related type of measure validation. However resulting population prevalence estimates varied substantially in size and demographic profile depending on the type of SDQ scoring method used.

These studies offer some reassurance that the extended and shorter SDQ versions are identifying key populations of interest (i.e. children who may have mental disorders). However, the method recommended for validating brief mental health screening measures is

administering the instrument of interest and gold standard clinical assessments in close succession so the brief measure can be calibrated to the clinical assessment for maximum concordance (Kessler et al., 2004). This approach was used to estimate the prevalence of adult serious mental illness (SMI) in the National Survey on Drug Use and Health (NSDUH) (Aldworth et al., 2010; Kessler et al., 2003). Similar work has been conducted with adolescents. In the National Comorbidity Survey- Adolescent Supplement (NCS-A), Kessler et al. (2006) examined associations between the five-symptom item SDQ and Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS) interviews (Kaufman et al., 1997). Observed K-SADS prevalence estimates of adolescent mental disorders with serious impairment were reproduced with little model-based bias (6.1% in the five-symptom (no impact) item SDQ versus 4.8% in the K-SADS) and with good individual-level concordance ( $AUC^* = .85$ ) (Kessler et al., 2006). The researchers concluded that the best scoring approach was to sum responses into a total score and use a cut point of 6 to estimate prevalence of mental disorders with serious impairment.

Screening measures are typically used to identify cases in need of further clinical assessment; consequently, these scales are typically short, easy-to-score and designed to be clinically relevant. To be most helpful screening scales need to be sensitive enough to identify the majority of relevant cases and also have high positive predictive value (PPV), or a low false positive rate (Glover & Albers, 2006; Kessler, Andrews, Colpe, Hiripi, & Zaslavsky, 2002). Several short screening scales exist to examine specific disorders in children, but far fewer short, comprehensive screeners are available to identify a more broad risk for any mental disorder. This article examines the utility of the very short global 6-item SDQ included in the NHIS for constructing models that can be used to predict the prevalence of any impairing child mental disorder as measured by the Child and Adolescent or Preschool-age Psychiatric Assessments (CAPA or PAPA). Using similar methods to those described by Kessler et al. (2003) and Aldworth et al. (2010), this study expands upon the NCS-A findings (Kessler et al., 2006) to examine prediction models for childhood mental disorders across a broad age range of children and adolescents included in the NHIS. We compare and contrast four predictive models that mirror the variety of scoring methods previously used (Federal Interagency Forum on Child and Family Statistics, 2012; Pastor et al., 2012). We also examine how child age (4–11 versus 12–17 years) impacts prediction methods for any impairing childhood mental disorder.

## Methods

### Sample

This study uses data collected from a follow-up study to the NHIS (Aldworth et al., 2012). NHIS data are collected through in-person household interviews among a nationally representative sample of the civilian noninstitutionalized U. S. population. When children are present in a family, one sample child is selected. Parent participants for this study were recruited from the final three quarters of 2011 and first quarter of 2012 of NHIS. Only sample children whose parents completed the NHIS interview in English, provided complete

---

\*AUC=area under the receiver operating characteristic (ROC) curve.

contact information and SDQ responses, and indicated the child had no history of mental retardation (now referred to as “intellectual disability”), developmental delay, autism, or Down syndrome were eligible for participation. In the current study, parents were the only respondents for children 4–11; for children 12–17, parent/child pairs were required. Cases were selected based on the Neyman optimal allocation design; sampling strata were defined by 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. In addition, the oversampling of minorities in the NHIS sampling design was partially reversed by a race/ethnicity adjustment factor applied to the design in order to control the variability in analysis weights.

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%), including 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.) Nonresponse and poststratification adjustments were applied to the weights to lessen bias due to nonresponse or the lack of population coverage meeting eligibility requirements.

Respondent characteristics for the NHIS main sample and this study sample are shown in Table 1. Demographic characteristics of the study sample based on the final study sample weights are identical to those of the main NHIS sample because both were poststratified to the same US Census-based totals. The SDQ weighted mean scores are also similar across the study and NHIS samples.

## Measures

**Strengths and Difficulties Questionnaire (SDQ)**—For the NHIS, parents report on children's mental health by responding to the 6-item SDQ (Goodman, 2001). Parents are asked the degree to which 5 items describe the sample child by responding “not true,” “somewhat true,” or “certainly true” to each item (eg, “he is often unhappy, depressed, or tearful”). Each item is scored 0–2; the total SDQ score sums the 5 items (range 0–10). These 5 items are supplemented by 1 “impact” item that assesses functional impairment: “Overall, do you think that [child] has difficulties in any of the following areas: emotions, concentration, behavior, or being able to get along with other people?” The parent is asked to respond “no” or “yes” to minor, definite, or severe difficulties, yielding a score of 0–3. This 6-item SDQ was developed in consultation with the scale developer to fit within the time constraints of the NHIS (U.S. Department of Health and Human Services/Centers for Disease Control and Prevention, 2003). Kessler and colleagues (2006) examined the parent-reported 25- and 5-item SDQ measures in a model predicting clinical interview results from the K-SADS (Kaufman et al., 1997). High item-total correlations (.64–.76) were found between the single items selected from the five SDQ subscales and scores from the full subscales included in the 25-item version (Kessler et al., 2006).

**Clinical Interview**—The clinical interview used in this study for children aged 8–17 was the electronic, shortened Child and Adolescent Psychiatric Assessment (CAPA) (Angold and Costello, 2000). The CAPA is a semistructured interview recommended for use with

children aged 9–18. It is based on the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV) (American Psychiatric Association, 1994) and is administered by trained lay interviewers. In the CAPA, respondents report symptoms for lifetime and current (past 3 months) time periods. The shortened CAPA instruments include five modules to assess anxiety disorders, mood disorders, attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), and conduct disorder (CD). The CAPA has established clinical validity and test-retest reliability across various mental disorder diagnoses ranges from kappa = 0.55 to 1.0 (Angold & Costello, 1995, 2000).

The clinical interview for children aged 4–7 was the electronic Preschool Age Psychiatric Assessment (E-PAPA) (Egger et al., 2006), the CAPA younger-age companion. It was developed for parents of children aged 2–5, although it has been used for children as old as 7. The test-retest reliability of the PAPA has been examined in a pediatric clinical sample with results comparable to diagnostic interviews with parents with older children (Egger et al., 2006). A shortened five-module version (anxiety, mood, ODD, ADHD, and CD) was developed specifically for use in this study.

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. “Partial incapacity” refers to a notable reduction of function. “Severe incapacity” refers to a complete, or almost complete, inability to function in a particular area. Two measures were derived from the CAPA and PAPA clinical interviews: mental disorder with impairment (MDI) and mental disorder with severe impairment (MDSI). MDI indicates a presence of any disorder assessed from the five modules with at least one partial or severe impairment rating; MDSI indicates the presence of any assessed disorder accompanied by one or more severe impairment rating. For children aged 12–17, MDI or MDSI was based on disorder and impairment information from both parent and child interviews.

**Data Collection**—Data collection occurred from February to August 2012. Verbal parent consent and adolescent assent were obtained from all respondents. To maintain a consistent reporting period between the clinical interview and screening measure, the 6-item SDQ was readministered at the start of the clinical interview. Parent respondents of children aged 4–7 were administered the PAPA (Egger et al., 2006). Parents of children aged 8–17 and children aged 12–17 were administered the CAPA (Angold and Costello, 2000). Interviews were conducted by telephone by trained lay interviewers using tablet computers. Interviews were recorded for quality review. Parent interviews were conducted first. Parents were provided a \$25 remuneration; children aged 12–17 received a \$25 gift card. Recruitment and consent procedures were approved by the contracting organization's Institutional Review Board.

**Statistical Analysis**—For each of the two age groups (4–11 and 12–17), weighted logistic regression models were applied to the study sample data in which the response variable was either CAPA/PAPA-based MDI or MDSI (positive or negative), and the explanatory variables included various formulations of the five symptom/one impact item SDQ scores. The study sample final analysis weights were used in the models.

Receiver operating characteristic (ROC) analyses were applied to each of the models to identify an appropriate cut point to dichotomize the predicted probability of MDI based on the model, for each of the two age groups. A cut point on the SDQ was needed so that prevalence estimates based on the SDQ were as close as possible to those based on the CAPA or PAPA. With a coarse instrument such as the SDQ with very few levels, it is not always possible to do this exactly, especially when dealing with weighted data. Therefore, any differences that result between the SDQ and CAPA estimates (positive or negative) represent the model-based bias. If a child's predicted probability of MDI was greater than or equal to the cut point, then that child was predicted to be MDI positive; otherwise, he or she was predicted to be MDI negative. The cut point was selected to (approximately) equalize the weighted false positive (ie, predicted MDI positive, but CAPA/PAPA-based MDI negative) and false negative (ie, predicted MDI negative, but CAPA/PAPA-based MDI positive) counts. This is equivalent to minimizing the model-based bias in prevalence estimates of MDI based on CAPA/PAPA determinations versus those based on predictions derived from the model and cut point. An identical approach was used to analyze MDSI.

## Results

### Models Using the SDQ Impact Item Only

As in *America's Children* (Federal Interagency Forum on Child and Family Statistics, 2012), the four-level SDQ impact item can be collapsed into a two-level variable to describe none or minor vs. definite or severe difficulties. This variable has only one (informative<sup>†</sup>) discriminating cut point; consequently, this is the only way it can predict childhood mental disorder (ie, it predicts both MDI and MDSI with the same cut point).

In contrast, models based on the four-level impact item variable have three possible informative cut points and hence the capacity to distinguish between MDI and MDSI prediction. Results of an ROC analysis applied to the four-level impact-item model, for both MDI and MDSI and for each age group, are given in Table 2, (Model One). Because the four-level impact item can score 0, 1, 2, or 3, a cut point of  $x$  indicates that a score of  $x$  or greater will result in a positive MDI or MDSI prediction. Note that a cut point of 2 for the four-level impact item is equivalent to a cut point of 1 for the dichotomized impact item. Therefore, the four rows of Table 2 in which the cut point is 2 exactly describe the ROC statistics of the dichotomized impact item; note that for these rows the cut point of 2 and modeled estimate are the same for MDI and MDSI (5.15 percent for children aged 4–11 and 7.32 percent for children aged 12–17). In the case of MDI prediction, the negative bias<sup>‡</sup> is large (–16.40 percentage points for children aged 4–11 and –13.68 percentage points for children aged 12–17). In contrast, models based on the four-level impact item are able to distinguish between MDI and MDSI prediction. Even so, the large positive bias associated with MDI prediction using a cut point of 1 (15.33 percentage points for children aged 4–11 and 9.20 percentage points for children aged 12–17), suggests that even the four-level impact item is graduated too coarsely to accurately predict MDI. The negative bias associated

<sup>†</sup>A cut point that assigns all cases as positive or all as negative is not considered informative.

<sup>‡</sup>Bias is the difference between modeled and P/CAPA estimates.



with MDSI prediction is much smaller (−3.67 percentage points for children aged 4–11 and −2.67 percentage points for children aged 12–17).

### Models Using the SDQ Score Only

Results of an ROC analysis applied to models with the 5-item (no impact item) SDQ score (hereafter called “SDQ Score”) as a predictor (Kessler et al., 2006; Pastor et al., 2012) appear in Table 2 (Model Two). Because the 5-item SDQ score has 11 levels (i.e., 0–10), it has a finer discriminating capacity than the 4-level impact item; this shows in the comparatively smaller levels of absolute bias in Table 2. Bias reduction is particularly noticeable in the case of MDSI prediction (0.50 percentage points for children aged 4–11 and 1.05 percentage points for children aged 12–17). However, for the 4–11 age group, sensitivity and AUCd (ie, AUC based on the SDQ score dichotomized by the cut point) are very small (sensitivity = 0.237; AUCd = 0.579). This indicates that although the dichotomized SDQ score results in a bias reduction (in comparison with the impact item), it is not an accurate method to predict MDSI for this age group.

### Models Based on SDQ Score and Impact Item

In the models including both the SDQ score and the impact item as predictors bias values are similar or slightly reduced relative to those from the SDQ score model (see Table 2, Model Three); for the 4–11 year age group, sensitivity and AUCd are substantially increased (sensitivity = 0.476; AUCd = 0.717; Table 2). ROC statistics of the other three cases for this model are similar or better (sensitivity = 0.579–0.668; AUCd = 0.715–0.789). These results suggest that the impact item added as a predictor improves the prediction of MDI or MDSI for both age groups.

Because this model has two independent predictors, cut points need to be defined with respect to both the impact item score and SDQ score simultaneously (Table 3). For example, for the 4–11 age group, a combination of an impact item score of 0 and an SDQ score  $\geq 3$  would be required to positively predict MDI, but in the case of MDSI prediction, no SDQ score cutpoint exists in combination with an impact item score of 0 to positively predict MDSI (i.e., even an SDQ score of 10 would not be sufficient to positively predict MDSI). If the impact item score is 1, then an SDQ score  $\geq 3$  would be required to positively predict MDI, but an SDQ score  $\geq 5$  would be required to positively predict MDSI. Table 3 shows some inconsistency in the MDI versus MDSI models for the 4–11 age group: if the impact item score is 2 or 3, then a larger cut point with respect to the SDQ score would be required to positively predict MDI than MDSI. Also of interest is the fact that the cut points vary by age group; in other words, a positive MDI or MDSI prediction depends both on impact item score and SDQ score and on age group.

### Models Based on Subset of Five Individual SDQ Items and Impact Item

To further reduce the absolute bias of modeled prevalence estimates, the five individual SDQ item scores and the impact item were included as predictors in the same model. These results are found in Table 2 (Model Four). Exact cut points are not provided in the table; however, for illustration purposes, a sample equation for calculating MDI among 4–11 year olds is included in Table 2, footnote 9. In these models, some of the SDQ items had negative

regression coefficients (ie, an increase in the item score would result in a lower probability of MDI or MDSI prediction); consequently, these terms were dropped from those models. One item in particular (“gets along better with adults than with other children”) behaved like this in all models, except in the case of MDSI prediction for the 12–17 age group. Another item (“has many worries”) was dropped from the individual item model for MDSI prediction for the 4–11 age group. Some ROC statistics show a reduction in the bias of the modeled estimates, but others remain about the same (Table 2). Although not displayed, these similar models without the impact item score as a predictor performed more poorly, indicating the importance of including the impact item in the models.

## Conclusion

This study investigates methods for using the 6-item SDQ included in the NHIS to estimate the prevalence of any impairing or severely impairing childhood mental disorder. Results demonstrate the importance of using all available information from the 6-item SDQ in accurately estimating the prevalence of childhood mental disorders from the NHIS. The two strongest models were those based on: (1) SDQ score and impact item and (2) the subset of five individual SDQ items and impact item. These models maximized the use of the SDQ items which, in turn, yielded a finer gradation of scores, a reduction of model-based bias, and improved sensitivity and concordance than approaches based exclusively on the total SDQ score (Kessler et al., 2006; Pastor et al., 2012) or the impact item score alone (Federal Interagency Forum on Child and Family Statistics, 2012).

Study results varied by child age and mental disorder outcomes, but not consistently. The total SDQ score model (Kessler et al., 2006; Pastor et al., 2012) did not perform well in predicting the most seriously impairing mental disorders (MDSI) among children aged 4–11. This finding indicates the model used previously for adolescents aged 13 or older (Kessler et al., 2006) may not apply to younger children. The model that used both total SDQ score and impact items better predicted MDSI among children aged 4–11 (smaller absolute bias, 0.85) and showed an even smaller bias (0.14) in predicting a lower threshold of MDI among children aged 12–17. But this model still showed somewhat large biases in predicting MDI for children aged 4–11 and MDSI for children aged 12–17 (4–11 years MDI bias = 3.63; 12–17 years MDSI bias = -2.06). The model that included some individual SDQ item scores instead of the total SDQ score (in addition to the impact item) showed the least fluctuation in terms of bias by child age and mental disorder outcome; but the most fluctuation in the items retained in the models for each age group. These models consistently showed lower absolute bias (i.e., smallest absolute bias = 0.03 and largest absolute bias = 0.43) than other models. Findings demonstrate the importance of considering both child age and outcome of interest in model selection. Models built upon adolescent samples may not function similarly with young children, and models that predict highly impairing disorders may not function similarly in predicting any mental disorder.

Data from the NHIS are publicly available from the NHIS website. The analytic models developed in this study for estimating MDI and MDSI across two age groups offer preliminary guidance for NHIS users on how to improve upon previous SDQ scoring methods (Federal Interagency Forum on Child and Family Statistics, 2012; Kessler et al.,



2006; Pastor et al., 2012). As Table 3 illustrates, cut points that depend upon the interaction of child age and outcome of interest can be reliably used when estimating the prevalence of MDI and MDSI. The online survey description document that accompanies the annual NHIS data release will include an appendix that gives further detail for users.

This study's results could be strengthened by future research. First, like many surveys, the NHIS uses a parent informant approach to child health data collection. The ability to predict childhood mental disorders within a national survey might improve with the addition of a child self-reported screening tool. Second, the response rate of the study and sample sizes were substantially lower than desired; these should be expanded in future work. Third, the study's selection of a "gold standard" psychiatric assessment may yield different results from those generated by a different clinical assessment. This study used a shortened 5-module CAPA/PAPA diagnostic interview; further studies should validate these results with a full-length clinical interview. CAPA results have been shown to be particularly powerful at eliminating false negative cases, but may yield lower overarching estimates of any mental disorders than a clinician-administered interview (Angold, Erkanli, Copeland, Goodman, Fisher, & Costello, 2012). Consequently, these results should also be replicated using a clinician-administered gold-standard interview. Finally, the 6-item SDQ included in the NHIS could benefit from further psychometric testing and refinement. While recommended models 3 and 4 from this study showed moderate concordance values (AUCd .71-.81), item refinement is needed to further strengthen concordance with clinical interview results. Furthermore, in the individual SDQ item models, some terms had negative regression coefficients and hence were excluded. Examining the ability of the 6-item SDQ to estimate the prevalence of individual mental disorders was beyond the scope of the current study; however, this too could be examined in future work. This study was the first to examine the ability of the 6-item SDQ to predict the prevalence of any impairing mental disorder among children 4 years and older as well as adolescents. The NHIS is the only national survey that annually includes an instrument to assess mental health broadly across a wide age range of children. Consequently, this research represents a necessary first step toward the goal of producing national prevalence estimates of any childhood mental disorder and associated impairment.

## Acknowledgments

The views expressed in this manuscript do not necessarily represent the views of the National Institutes of Health, the National Center for Health Statistics (NCHS) or the Federal Government. This research was supported by National Institute of Mental Health through contract 200-2009-F-32679 with the NCHS within the Centers for Disease Control and Prevention. We acknowledge reviews by Joe Gfroerer, Jonaki Bose, and Sarra Hedden from the Substance Abuse and Mental Health Services Administration and Patricia Pastor, Marcie Cynamon, Stephen Blumberg, and Jennifer Madans from the National Center for Health Statistics. At RTI International, Anne Gering and Judith Cannada provided editorial assistance and report production. The statistical expert was Jeremy Aldworth.

## References

- Aldworth J, Colpe LJ, Gfroerer JC, Novak SP, Chromy JR, Barker PR, Barnett-Walker K, Karg RS, Morton KB, Spagnola K. The National Survey on Drug Use and Health Mental Health Surveillance Study: calibration analysis. *Int J Methods Psychiatr Res.* 2010; 19 Suppl 1:61–87.10.1002/mpr.312 [PubMed: 20527006]

- Aldworth, J.; Ringeisen, H.; Morgan, K. Presented to the National Center for Health Statistics, Centers for Disease Control and Prevention. 2012. Developing model-based SED estimates in large-scale surveys: Pilot study results.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Washington, DC: Author; 1994.
- Angold A, Costello EJ. A test-retest reliability study of child-reported psychiatric symptoms and diagnoses using the Child and Adolescent Psychiatric Assessment (CAPA-C). *Psychological Medicine*. 1995; 25:755–762. [PubMed: 7480452]
- Angold A, Costello EJ. The Child and Adolescent Psychiatric Assessment (CAPA). *J Am Acad Child Adolesc Psychiatry*. 2000; 39(1):39–48.10.1097/00004583-200001000-00015 [PubMed: 10638066]
- Angold A, Erkanli A, Copeland W, Goodman R, Fisher PW, Costello EJ. Psychiatric diagnostic interviews for children and adolescents: A comparative study. *J Am Acad Child Adolesc Psychiatry*. 2012; 51(5):506–517. [PubMed: 22525957]
- Bourdon KH, Goodman R, Rae DS, Simpson G, Koretz DS. The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. *J Am Acad Child Adolesc Psychiatry*. 2005; 44(6):557–564.10.1097/01.chi.0000159157.57075.c8 [PubMed: 15908838]
- Egger HL, Erkanli A, Keeler G, Potts E, Walter BK, Angold A. Test-Retest Reliability of the Preschool Age Psychiatric Assessment (PAPA). *J Am Acad Child Adolesc Psychiatry*. 2006; 45(5): 538–549.10.1097/01.chi.0000205705.71194.b8 [PubMed: 16601400]
- Federal Interagency Forum on Child and Family Statistics. America's children in brief: Key national indicators of well-being, 2012. Washington, DC: U.S. Government Printing Office; 2012.
- Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry*. 2001; 40(11):1337–1345.10.1097/00004583-200111000-00015 [PubMed: 11699809]
- Glover TA, Albers CA. Considerations for evaluating universal screening instruments. *J School Psychol*. 2006; 45:117–135.10.1016/j.jsp.2006.05.005
- Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N. 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*. 1997; 36(7): 980–988.10.1097/00004583-199707000-00021 [PubMed: 9204677]
- Kessler RC, Abelson J, Demler O, Escobar JI, Gibbon M, Guyer ME, Howes MJ, Jin R, Vega WA, Walters EE, Wang P, Zaslavsky A, Zheng H. Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMHCIDI). *Int J Methods Psychiatr Res*. 2004; 13(2):122–139. [PubMed: 15297907]
- Kessler RC, Andrews G, Colpe LJ, Hiripi E, Zaslavsky AM. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med*. 2002; 32:959–976. doi:10.1017/S0033291702006074. [PubMed: 12214795]
- Kessler RC, Avenevoli S, Costello EJ, Georgiades K, Green JG, Gruber MJ, He JP, Koretz D, McLaughlin KA, Petukhova M, Sampson NA, Zaslavsky AM, Merikangas KR. Prevalence, persistence, and sociodemographic correlates of DSM-IV disorders in the National Comorbidity Survey Replication Adolescent Supplement. *Arch Gen Psychiatry*. 2012; 69(4):372–380.10.1001/archgenpsychiatry.2011.160 [PubMed: 22147808]
- Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, Howes MJ, Normand SL, Manderscheid RW, Walters EE, Zaslavsky AM. Screening for serious mental illness in the general population. *Arch Gen Psychiatry*. 2003; 60(2):184–189. [PubMed: 12578436]
- Kessler, RC.; Gruber, M.; Sampson, NA. Presented to the Centers for Disease Control and Prevention. Boston, MA: Harvard Medical School; 2006. Validation studies of mental health indices in the National Health Interview Survey, Final Report and Addendum.
- Merikangas KR, He JP, Brody D, Fisher PW, Bourdon K, Koretz DS. Prevalence and treatment of mental disorders among US children in the 2001–2004 NHANES. *Pediatrics*. 2010; 125(1):75–81.10.1542/peds.2008-2598 [PubMed: 20008426]

- Pastor PN, Reuben CA, Duran CR. Identifying emotional and behavioral problems in children aged 4–17 years: United States, 2001–2007. *Natl Health Stat Report*. 2012; (48):1–17. [PubMed: 22737946]
- U.S. Department of Health and Human Services. Report of the Surgeon General's conference on children's mental health: A national action agenda. Washington, DC: U.S. Department of Health and Human Services; 2000.
- U.S. Department of Health and Human Services/Centers for Disease Control and Prevention. 2002 National Health Interview Survey (NHIS) public use data release: NHIS survey description. Hyattsville, MD; Division of Health Interview Statistics, National Center for Health Statistics; 2003.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1**  
**Child Population Estimates (Percentage Distributions)**

Variable	NHIS Main Sample (NHIS Sample Child Weights)	Final Study Sample (Design Weights <sup>1</sup> )	Final Study Sample (Final Sample Weights <sup>2</sup> )
<b>Age Group</b>			
4–11	57.2	66.2	57.2
12–17	42.8	33.8	42.8
<b>Gender</b>			
Male	51.2	47.4	51.2
Female	48.8	52.6	48.8
<b>Race/Ethnicity</b>			
Non-Hispanic White	54.2	55.8	54.2
Non-Hispanic Black	13.1	19.9	13.1
Non-Hispanic Other	9.9	16.4	9.9
Hispanic	22.8	7.9	22.8
<b>NHIS 5-item SDQ Score<sup>3</sup></b>			
0	32.0	30.4	31.6
1	23.6	23.1	20.1
2	20.0	17.4	19.5
3	10.3	14.3	13.0
4	6.1	7.1	8.3
5	3.7	4.0	3.9
6	2.4	2.0	2.1
7	0.9	0.6	0.5
8	0.6	0.6	0.5
9	0.2	0.4	0.5
10	0.0	0.0	0.0
<b>NHIS SDQ Impact Item Score<sup>3</sup></b>			
0	80.1	81.4	78.6
1	14.4	15.4	17.0
2	4.1	2.2	3.1
3	1.3	1.1	1.3

NHIS = National Health Interview Survey; SDQ = Strengths and Difficulties Questionnaire.

<sup>1</sup> Design weight is the NHIS sample child weight multiplied by reciprocal of study sample selection probability.

<sup>2</sup> Final sample weight is the design weight modified by nonresponse adjustments and post-stratification adjustments to match the NHIS control totals by age group, gender, and race/ethnicity.

<sup>3</sup> Cases with missing SDQ score and impact item score values were excluded from this table.

Table 2

Receiver Operating Characteristic Model Statistics

Model One = Four-Level Impact Item (0 to 3) Predictor												
Age Group	Variable	Cut Point <sup>l</sup>	P/CAPA Estimate <sup>2</sup>	Modeled Estimate <sup>2</sup>	Bias <sup>2</sup>	FP <sup>3</sup> (Percent)	FN <sup>3</sup> (Percent)	Sens <sup>4</sup>	Spec <sup>4</sup>	PPV <sup>5</sup>	NPV <sup>5</sup>	AUCc <sup>6</sup>
4-11	MDI	1	21.55	36.88	15.33	25.09	9.76	0.547	0.680	0.320	0.845	0.614
4-11	MDI	2	21.55	5.15	-16.40	1.02	17.43	0.191	0.987	0.801	0.816	0.589
4-11	MDSI	2	8.81	5.15	-3.67	1.35	5.02	0.430	0.985	0.737	0.947	0.708
12-17	MDI	1	21.00	30.20	9.20	11.32	2.12	0.899	0.857	0.625	0.970	0.878
12-17	MDI	2	21.00	7.32	-13.68	1.78	15.46	0.264	0.977	0.756	0.833	0.621
12-17	MDSI	2	9.99	7.32	-2.67	1.93	4.60	0.539	0.979	0.736	0.950	0.759
Model Two = SDQ Score (0 to 10) Predictor												
Age Group	Variable	Cut Point <sup>l</sup>	P/CAPA Estimate <sup>2</sup>	Modeled Estimate <sup>2</sup>	Bias <sup>2</sup>	FP <sup>3</sup> (Percent)	FN <sup>3</sup> (Percent)	Sens <sup>4</sup>	Spec <sup>4</sup>	PPV <sup>5</sup>	NPV <sup>5</sup>	AUCc <sup>6</sup>
4-11	MDI	3	21.55	25.07	3.52	12.37	8.85	0.589	0.842	0.507	0.882	0.716
4-11	MDSI	4	8.81	9.32	0.50	7.23	6.73	0.237	0.921	0.224	0.926	0.579
12-17	MDI	4	21.00	26.13	5.13	13.54	8.41	0.599	0.829	0.482	0.886	0.714
12-17	MDSI	5	9.99	11.04	1.05	5.41	4.35	0.564	0.940	0.510	0.951	0.752
Model Three=SDQ Score (0 to 10) and Impact Item (0 to 3) Predictors												
Age Group	Variable	Cut Point <sup>l</sup>	P/CAPA Estimate <sup>2</sup>	Modeled Estimate <sup>2</sup>	Bias <sup>2</sup>	FP <sup>3</sup> (Percent)	FN <sup>3</sup> (Percent)	Sens <sup>4</sup>	Spec <sup>4</sup>	PPV <sup>5</sup>	NPV <sup>5</sup>	AUCc <sup>6</sup>
4-11	MDI		21.55	25.18	3.63	12.47	8.85	0.589	0.841	0.505	0.882	0.715
4-11	MDSI		8.81	7.97	-0.85	3.77	4.62	0.476	0.959	0.526	0.950	0.717
12-17	MDI		21.00	21.14	0.14	7.11	6.97	0.668	0.910	0.664	0.912	0.789
12-17	MDSI		9.99	7.93	-2.06	2.15	4.21	0.579	0.976	0.729	0.954	0.777
Model Four=Subset of up to 5 Individual SDQ Item Scores (0 to 2) and Impact Item (0 to 3) Predictors												
Age Group	Variable	Cut Point <sup>8</sup>	P/CAPA Estimate <sup>2</sup>	Modeled Estimate <sup>2</sup>	Bias <sup>2</sup>	FP <sup>3</sup> (Percent)	FN <sup>3</sup> (Percent)	Sens <sup>4</sup>	Spec <sup>4</sup>	PPV <sup>5</sup>	NPV <sup>5</sup>	AUCc <sup>6</sup>
4-11	MDI <sup>9</sup>		21.55	21.12	-0.43	7.59	8.01	0.628	0.903	0.641	0.898	0.766
4-11	MDSI		8.81	8.92	0.11	4.77	4.66	0.471	0.948	0.466	0.949	0.710
12-17	MDI		21.00	21.24	0.24	7.84	7.60	0.638	0.901	0.631	0.904	0.770
12-17	MDSI		9.99	10.02	0.03	3.34	3.31	0.669	0.963	0.667	0.963	0.816

AUC=area under ROC curve (continuous predictor); AUCd=area under ROC curve (dichotomous predictor); FN=false positive; MDI=mental disorder with impairment; MDSI=mental disorder with severe impairment; NPV=negative predictive value; P/CAPA=preschool age psychiatric assessment (PAPA) or child and adolescent psychiatric assessment (CAPA); PPV=positive predictive value; ROC=receiver operating characteristic; Sens=sensitivity; Spec=specificity.

- 1 For models with a single predictor, a cut point of  $x$  indicates that a score  $x$  will result in a positive MDI or MDSI prediction; for models with multiple predictors, the cut point is not given (see Table 2, footnotes 7 and 8).
- 2 P/CAPA Estimate and Modeled Estimate refer to prevalence estimates of MDI or MDSI based on P/CAPA determinations and model predictions, respectively. Bias is the difference between modeled and P/CAPA estimates.
- 3 FP and FN are percentages of weighted false positive and false negative counts, respectively.
- 4 Sens (ie, sensitivity) is the weighted proportion of P/CAPA MDI or MDSI positive cases also predicted MDI or MDSI positive (ie, positive “hit” rate). Spec (ie, specificity) is the weighted proportion of P/CAPA MDI or MDSI negative cases also predicted MDI or MDSI negative (ie, 1 - “false alarm” rate).
- 5 PPV is the weighted proportion of predicted MDI or MDSI positive cases that are also P/CAPA MDI OR MDSI positive. NPV is the weighted proportion of predicted MDI or MDSI negative cases that are also P/CAPA MDI or MDSI negative.
- 6 AUCc is the concordance between the model (with continuous predictor) and P/CAPA MDI or MDSI, and AUCd is the concordance between the model (with predictor dichotomized by the cut point) and P/CAPA MDI or MDSI; AUCd = the average of sensitivity and specificity.
- 7 See Table 3 for cut points determined for combination of impact item score and SDQ score.
- 8 Cut point not given, because it would involve multiway tables or a detailed model description and cut point in terms of predicted probabilities.
- 9 Example formula: MDI (4-11 years) =  $\frac{1}{1 + \exp[-(-3.7309 + 2.8261 * \text{Obed} + 1.1503 * \text{Worry} + 2.0109 * \text{Unhappy} + 0.2049 * \text{Atten} + 0.6565 * \text{Impact}]}$  0.46940 where Obed(iance), Worry, Unhappy, Adults, and Atten(ion) refer to the 5 items used in the abbreviated SDQ§ and impact refers to a sixth SDQ item. Other formulas available upon request.



**Table 3**  
**Cut Points Based on Example 3: Impact Item Score and SDQ Score**

Age Group	Impact Item Score	5-item SDQ Score Cut Point <sup>I</sup>	
		MDI	MDSI
4-11	0	3	*
	1	3	5
	2	2	0
	3	1	0
12-17	0	7	8
	1	3	6
	2	0	3
	3	0	1

MDI = mental disorder with impairment, MDSI = mental disorder with severe impairment, SDQ = Strengths and Difficulties Questionnaire.

\* No SDQ score cutpoint combined with Impact score = 0 predicted MDSI status for the 4-11 age group

<sup>I</sup> A cut point of  $x$  indicates that an SDQ score  $\geq x$  will result in a positive MDI or MDSI prediction.