The Baby Pediatric Symptom Checklist: Development and Initial Validation of a New Social/Emotional Screening Instrument for Very Young Children

R. Christopher Sheldrick, PhD; Brandi S. Henson, PsyD; Emily N. Neger, BA; Shela Merchant, MA; J. Michael Murphy, EdD; Ellen C. Perrin, MD

From the Division of Developmental-Behavioral Pediatrics, Floating Hospital for Children, Tufts Medical Center, Boston, Mass (Dr Sheldrick, Dr Henson, Ms Neger, Ms Merchant, and Dr Perrin); and Department of Psychiatry, Massachusetts General Hospital, Boston, Mass (Dr Murphy)
The authors have no conflicts of interest to disclose.
Research support for the development and validation of the Baby Pediatric Symptom Checklist was provided by The Commonwealth Fund & National Institutes of Health grant KM1CA156726.
Address correspondence to R. Christopher Sheldrick, PhD, 800 Washington Street, Box 854, Boston, Massachusetts 02111 (e-mail: rsheldrick@tuftsmedicalcenter.org).
Received for publication June 26, 2012; accepted August 15, 2012.

ABSTRACT

OBJECTIVE: To develop and validate the Baby Pediatric Symptom Checklist (BPSC), a brief social/emotional screening instrument for children less than 18 months. The BPSC is modeled after the Pediatric Symptom Checklist (PSC) and is part of the Survey of Wellbeing of Young Children, a comprehensive, freely available screening instrument designed for use in pediatric primary care.

METHOD: BPSC items were developed by a team of experts who reviewed existing assessment instruments and relevant research literature. Scale construction and initial validation were conducted with 205 families from pediatric primary care sites and 54 families from referral clinics. A replication sample of 146 additional families were enrolled from an independent set of primary care practices.

RESULTS: Exploratory factor analysis revealed 3 dimensions of the BPSC: irritability, inflexibility, and difficulty with routines. Factor structure was confirmed in the replication sample. Retest reliability and internal reliability were adequate (intraclass correlation coefficient >0.70) across subscales, with the exception of the "irritability" subscale's internal reliability in the replication sample. Construct validity of the "irritability" and the "difficulty with routines" subscales is supported by correlations with the Parenting Stress Index and the Ages & Stages Questionnaire: Social/Emotional, but the "inflexibility" subscale seems to be distinct from performance on these instruments. Tests of differential item functioning revealed no significant effects for race/ethnicity, child gender, parent education, or family income. Age-based normative data were calculated for each subscale.

CONCLUSION: The BPSC assesses 3 domains of behavior for very young children and shows promise as a social/emotional screening instrument for pediatric primary care.

KEYWORDS: social; emotional; behavioral; screening; pediatrics

ACADEMIC PEDIATRICS 2013;13:72–80

WHAT'S NEW

The BPSC is a brief social/emotional screening instrument designed for use in pediatrics with children under 18 months of age. It is easy to administer and score and is freely available. Initial investigation suggests that it has sound psychometric properties and effectively identifies social/emotional problems relevant to very young children.

APPROXIMATELY 10% TO 15% of 1-year-old and 2-year-old children experience social/emotional problems that cause significant problems for both the child and family.[1] Additionally, ample evidence suggests that early behavior patterns can predict the later emergence of social/emotional disorders as well as certain medical outcomes. For example, longitudinal studies have found that infants high in reactivity or behavioral inhibition at 4 months of age continue to react strongly and show symptoms of anxiety during the first 2 years of life and later into childhood.[2,3] Similarly, infants with temperaments labeled as "difficult" are 4 times as likely to have trouble adjusting in preschool and school[4] and are more likely to exhibit aggression and rule-breaking behaviors.[5] Early behavior patterns can also affect the quality of parent-infant interactions and are associated with both externalizing and internalizing symptoms later in childhood, especially if there is a mismatch between parenting style and child temperament.[6–8]

Medically, infants with "difficult" temperaments and high levels of negative reactivity have been found to be more accident-prone and to be more likely to experience a hospitalization later in childhood.[9] Highly emotional

infant temperament has been found to predict higher rates of childhood obesity[10] and children described as "difficult" in infancy are at significantly higher risk for tooth caries,[11] perhaps because parents have more difficulty getting them to comply with oral hygiene practices.

Despite the prevalence and long-term implications of early childhood behavioral difficulties, they are rarely detected, and fewer than 8% receive mental health services,[12] a much lower proportion than for older children.[13] There are several reasons for this discrepancy, including: (1) neither prominent child-focused psychosocial interventions (eg, cognitive-behavioral therapy) nor psychoactive medication is appropriate for very young children; (2) very few programs or therapists focus on infant mental health or relational psychotherapy with parents and young children; and (3) there are few validated behavioral screening or assessment instruments for infants.

Nevertheless, several beneficial interventions exist. Home-based family counseling for parents of infants with "difficult" temperaments have been found to reduce the incidence of psychiatric symptoms when children reach adolescence.[14] Home visiting programs have been shown to reduce the number of hospital visits and improve parental coping[15] and depression.[16] Counseling programs and parenting classes held in primary care and other settings have also proven to be effective interventions in the first years of life,[17] with education focused on providing families with alternative strategies for interacting with their children when a "temperamental mismatch" exists.[18] Early identification of emotional/behavioral problems in very young children can increase access to such services and may minimize later difficulties in school and social functioning.

Furthermore, because pediatric surveillance is by definition a longitudinal process,[19] assessing risk for behavioral disorders early may yield benefits later on. Assessment with a parent-completed screening instrument beginning at early ages helps parents learn that the pediatric office is an appropriate place to discuss child behavior and provides a longitudinal behavioral history for the physician to consider if faced with questions during subsequent visits. A large study by the Pediatric Research in Office Settings and the Ambulatory Sentinel Practice Network found that pediatricians were more likely to prescribe medications for a behavioral problem if concerns had already been noted in a previous visit.[20]

Screening programs designed to monitor social/emotional wellbeing have become increasingly prevalent in pediatrics because of national recommendations and assorted legal mandates,[21] yet many barriers still exist. Beyond the challenges of creating accurate instruments for use with infants, screeners must also be short, easy to read, simple to score, and inexpensive or free to use. The Pediatric Symptom Checklist (PSC) meets these criteria and has been well validated across a range of studies.[22–26] The PSC has become very popular as a screening instrument for children 4 years old and above in pediatric practice.[27,28] We worked with the creators of the PSC to develop similar instruments for younger children: the Baby Pediatric Symptom Checklist (BPSC) for children younger than 18 months and the Preschool Pediatric Symptom Checklist (PPSC) for children from 18 to 60 months.

The creation of these measures for younger children occurred as part of an ongoing project to develop a comprehensive surveillance instrument for children under 5 years of age, known as the Survey of Wellbeing of Young Children (SWYC). More information regarding the background and conceptualization of the SWYC is available in an earlier publication describing the PPSC[29] and on the website www.TheSWYC.org.

In this article, we describe the development and initial validation of the BPSC. Like the PSC, the BPSC is designed to maximize feasibility in clinical settings: it is brief, easy to score, and freely available.

METHODS

OVERVIEW

We created a list of possible items for the BPSC based on an extensive review of existing assessments and relevant research literature, as well as consultation with parents of young children and experts in child development. We enrolled 2 samples of parents to develop and pilot-test the BPSC: (1) a large original sample including parents from primary care sites and a small number from referral clinics (hereafter known as "original sample") and (2) an independent replication sample of parents from a different set of primary care pediatric practices (hereafter known as "replication sample"). Using data from the original sample, we conducted analyses to reduce the number of items and determine factor structure. Additionally, we conducted initial tests of internal reliability and construct validity. We then assessed the final abbreviated version of the BPSC with the replication sample. To establish retest reliability, approximately one third of the replication sample was asked to complete the BPSC a second time 3 to 4 weeks later. All studies were approved by the Institutional Review Board of Tufts Medical Center.

BPSC ITEM DEVELOPMENT AND DESCRIPTION

Our goal was to write questions that could be answered efficiently by parents from a range of educational and cultural backgrounds in the context of a pediatric waiting room. Thus, we attempted to write questions that were short, easy to read, and salient to parents. We began by identifying common questions and constructs across several parent-report measures designed for children under 18 months, including the Infant-Toddler Social and Emotional Assessment,[30] the Ages & Stages Questionnaire: Social/Emotional (ASQ:SE),[31] the Greenspan Social-Emotional Growth Chart,[32] the Behavioral Assessment of Baby's Emotional and Social Style,[33] and the Temperament and Atypical Behavior Scale.[34] In addition, we reviewed relevant literature on temperament and infant behavior and generated items based on our clinical experience.

Both the BPSC and PPSC are unique in that in addition to questions about child behavior, we have included questions that address parenting challenges. Inclusion of such questions is consistent with a transactional model in which

high levels of behavior problems in children increase parenting stress, whereas at the same time, increased levels of parenting stress contribute to a worsening of child behavior problems.[35] Past research has shown a strong correlation between parents' reports of child behavior problems and parenting stress,[36] and parental stress has been found to predict pediatric referrals.[37]

The initial list of BPSC items was reviewed by a group of 8 parents of young children and 11 experts in child development. Reviewers provided feedback regarding clarity, reading level, and relevance of items. This process resulted in 25 draft BPSC questions to be assessed in the original sample. These questions were screened for Flesch-Kincaid reading level. Items with reading levels over grade 6 were further reviewed and rewritten when possible. The average reading level of the final items was grade 2.1. For each item (eg, "does your child have a hard time calming down?" or "is your child fussy or irritable?"), response options included "not at all", "somewhat," and "very much."

Study Samples

For the original sample, parents of children younger than 18 months were enrolled from primary care practices (n = 205) and referral clinics (n = 54) in the greater Boston area. Primary care sites consisted of 4 urban practices and community health centers and 4 suburban practice groups. Referral sites consisted of 4 developmental-behavioral assessment clinics (including 2 neonatal intensive care units [NICUs] follow-up programs). For the replication sample (n = 146), parents were recruited from 6 unique primary care sites, including 3 suburban private practices and 3 urban health centers.

Enrollment for each stage occurred as follows:

Procedures

Enrollment procedures were identical to those used to develop the PPSC.[29] In primary care practices, research assistants approached parents in waiting rooms, described the study, and asked them if they would be interested in participating. In referral clinics (which had lower patient volumes), eligible parents were identified from health records and physicians mailed letters to parents describing the study, stating that a research assistant would call unless the parents indicated their wish not to be contacted (by calling a dedicated voice mail number).

The enrollment process for each procedure is depicted in Figure 1. Of potentially eligible families identified in waiting rooms, 90% enrolled in the study and 81% of enrolled parents completed all study protocols. Of potentially eligible families identified from medical records, 59% enrolled in the study and 86% of enrolled parents completed all study procedures.

Assessments

Because no single parent-report criterion measure of infant behavior is valid for children from birth through 18 months, we decided to administer several measures of constructs similar to what the BPSC is intended to assess. We chose 3 comparison instruments: (1) the ASQ:SE, a screening instrument intended to reflect social-emotional status among children starting at 3 months of age; (2) the Parenting Stress Index-Short Form (PSI/SF),[38] which consists of 3 subscales labeled "parental distress," "dysfunctional interactions," and "difficult child"; and (3) the 2-item Patient Health Questionnaire (PHQ-2),[39,40] a well-validated brief screening tool used to identify risk for depression among adults. We expected to find meaningful correlations between scores on the BPSC, the ASQ:SE, the "difficult child" subscale of the PSI/SF, and the PHQ-2.

At enrollment, parents were given a packet of questionnaires to complete about their child and mail back. For the original sample, the packets consisted of the 25 draft BPSC questions along with validated instruments including the age-specific form of the ASQ:SE, the PHQ-2, and demographic information. Three separate aged-based forms of the ASQ:SE were used to assess children 3 to 8 months, 9 to 14 months, and 15 to 18 months of age, respectively, based on published guidelines for the use of this instrument. The lengths of each form vary from 22 to 29 items. Assessments given to the replication sample were similar with 2 exceptions: (1) based on initial analyses (see below), a shorter form of the BPSC was administered that included 12 final items and (2) parents were asked to fill out the PSI.

Analyses

Four sets of analyses were conducted using Stata, version 12 (StataCorp, College Station, Tex) and Mplus, version 6.11 (Muthén & Muthén, Los Angeles, Calif): (1) construction of an abbreviated instrument; (2) factor structure and reliability; (3) concurrent validity, and (4) description of normative data.

1. Construction of an Abbreviated Instrument

To achieve the goal of creating an accurate screening instrument that is feasible for use in primary care, we first calculated descriptive statistics for responses to the 25 BPSC questions collected from the original sample. Then we calculated the frequency of each response category and of missing data, and eliminated items with ≥1% of missing data. We reasoned that more parents from the normative primary care setting would report an absence of each item ("not at all" response) than would report its presence ("somewhat" or "very much" responses); thus, we eliminated items that did not follow this pattern.

In order to further determine which items would be included in the final BPSC we constructed a latent variable model of remaining items using the original sample, comparing both bifactor and multifactor specifications.[41–43] In a bifactor model, every item loads onto a single general factor, and each item may also load on one or more additional factors.[44] To construct these subscales, we conducted exploratory factor analysis (EFA) within a confirmatory framework, choosing items that displayed the highest loadings that adequately represented the domains of interest, and that displayed the smallest degree of differential item functioning (DIF) with respect to enrollment site and demographic variables.

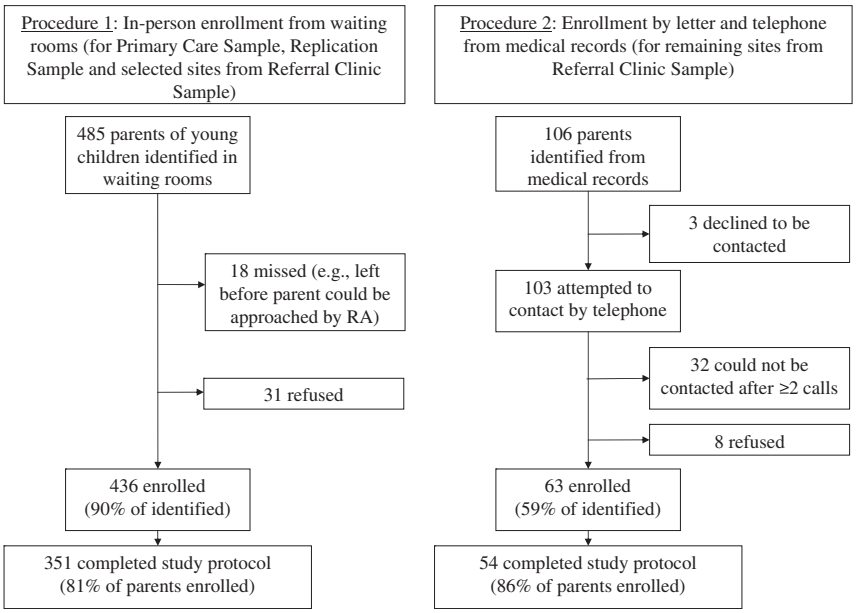

Figure 1. *Baby Pediatric Symptom Checklist (BPSC) enrollment.*

An item displays differential item functioning if, after controlling for underlying traits, responses differ between 2 population groups.

To evaluate the validity of interpreting the general factor as a primary dimension and choose a final model of factor structure, we compared the final bifactor model to a unidimensional model, in which all items load onto a single factor, and to a first-order multifactor model created based on a separate EFA. We also compared factor loadings and variance explained in the bifactor model between the general factor and the item-cluster factors. See Appendix-online only for further details regarding methods of scale construction.

2. Factor Structure and Reliability

After selecting a final set of items and constructing a final factor model, we tested its fit to the data in the original sample, and we also conducted a confirmatory factor analysis in the replication sample. Although items with relatively large DIF were eliminated during the process of creating an abbreviated scale, some degree of DIF could remain in the final set of items. To characterize the degree of DIF in our final model, DIF for each item in the final scale was expressed in terms of an odds ratio. To assess reliability, we calculated Cronbach's alpha for BPSC subscales in the original sample and again for the replication sample. Consistent with its statistical properties, Cronbach's alpha was interpreted as a lower bound of each subscale's reliability, rather than as an estimate of internal consistency.[45] In addition, 68 participants from the replication sample were asked to complete the BPSC a second time approximately 4 weeks later. Of these, 51 (75%) provided complete data. Retest reliability was calculated based on their answers using an intraclass correlation coefficient.

3. Concurrent Validity

To test concurrent validity, we scored each BPSC subscale and calculated Pearson correlations with the ASQ:SE and the PHQ-2 for both the original and replication samples. For the replication sample, we also calculated Pearson correlations between the BPSC and the PSI. Subscales were scored by summing across items, each of which was assigned 0 for a "not at all" response, 1 for a "somewhat" response, and 2 for a "very much" response. Because we used 3 different age-specific forms of the ASQ:SE in our study, scores were first standardized by form before correlations were calculated.

4. Normative Data

To facilitate interpretation, we analyzed normative data for the final version of the BPSC. Using quantile regression, cutoffs for the 50th, 70th, and 90th percentiles were calculated.

RESULTS

Characteristics for both the original and replication samples are reported in Table 1. For both samples, the majority of respondents were mothers (88% and 79%). Slightly more than half held a college degree or higher (54% and 56%). Family incomes varied, with just over a third of both samples making less than $50,000 per year (34% and 36%). Both samples were diverse with 35% and 38% reporting minority race or Hispanic ethnicity, respectively.

CREATION OF AN ABBREVIATED INSTRUMENT

We calculated descriptive statistics for each item. Responses for 1 item were missing for >1% of the sample. For an additional 2 items, "somewhat" responses were more common than "not at all" responses in the primary care sample, suggesting that the items assessed normative behaviors (eg, "moves around a lot" and "needs a lot of attention"). Items in both categories were dropped. We conducted an EFA in a confirmatory factor analysis framework in order to construct a preliminary factor model. Then we added binary variables for parent education, enrollment

Table 1. Sample Characteristics

	Original Sample		Replication Sample	
No. of patients	259		146	
Child male	131	51%	74	51%
Child Hispanic	36	14%	24	16%
Child race				
White	169	65%	90	62%
African American	24	9%	35	24%
Native American	4	2%	0	0%
Asian	27	10%	14	10%
Other/multiple races	11	4%	3	2%
Not indicated	9	3%	4	3%
Child age in months				
<6	81	31%	48	33%
7–12	95	37%	59	40%
13–18	83	32%	39	27%
Premature birth (<37 wks)	65	25%	14	10%
Public health insurance	50	19%	32	22%
Mother completed forms	228	88%	116	79%
Parent education				
<High school	18	7%	12	8%
High school diploma	57	22%	32	22%
Some college	40	15%	16	11%
College diploma	77	30%	35	24%
Advanced degree	61	24%	46	32%
Family income				
<$20,000	55	21%	34	23%
$20,000–49,999	33	13%	19	13%
$50,000–99,999	71	27%	34	23%
>$100,000	89	34%	55	38%
Not indicated	11	4%	4	3%
Site				
Primary care	205	79%	146	100%
Developmental follow-up clinic	54	21%	—	—

site, and child gender, age, race, and ethnicity to form a Multiple Indicator Multiple Cause model to identify items with large DIF. A final set of 12 items was selected by consensus based on their factor loadings, balance among domains of interest, and lack of DIF.

A 3-factor solution proved adequate in EFA, corresponded to our initial hypotheses, and was used as the basis of a bifactor and a multifactorial model. Most items in the bifactor model loaded more heavily on their individual factors rather than on the general factor, suggesting that the general factor is not an adequate representation of all items. The multifactorial model was more parsimonious and displayed adequate fit to the data; therefore, we chose this specification as our final model.

Factor Structure and Reliability

The final factor structure displayed adequate fit in both the original sample (root mean square errors of approximation = 0.046, Comparative Fit Index = 0.978, Tucker Lewis Index = 0.971), and the replication sample (root mean square errors of approximation = 0.037, Comparative Fit Index = 0.971, Tucker Lewis Index = 0.962). Final model structure and best estimates of standardized factor loadings and thresholds, calculated on all participants from primary care in either sample, are presented in Table 2. Three factors were interpretable, and we labeled these factors "irritability," "inflexibility" and "difficulty with routines." We tested each of the BPSC's 12 items for DIF with respect to 5 separate covariates, yielding a total of 60 separate tests. Among these, we identified 4 (6.7%) with odds ratios >2, indicating relatively large DIF with respect to child age. No DIF effects were found for parent education or child race, ethnicity or gender.

Cronbach's alpha values for each subscale are displayed in Table 3. Note that we also display Cronbach's alpha for other scales in our study, both for comparison and to facilitate interpretation of correlations from section 4 below. Cronbach's alpha—designed to estimate the lower bound of a scale's reliability[46]—was adequate across subscales, with the exception that the internal reliability of the "irritability" subscale fell below 0.70 in the replication sample. Retest reliability was adequate across subscales, with estimates of 0.70 for "irritability," 0.81 for "inflexibility," and 0.78 for "difficulty with routines."

Concurrent Validity

Sums were calculated for each BPSC subscale, yielding scores ranging from 0 to 8 for each of the 3 subscales. For each of the samples, correlations between the BPSC and the ASQ:SE and PHQ-2 are presented in Table 3. The "irritability" and the "difficulty with routines" subscales displayed moderate correlations with the ASQ:SE in both the original and the replication samples. Both also displayed correlations with the PHQ-2 in the original sample and the PSI Difficult Child Scale in the replication sample. The "inflexibility" subscale did not correlate with the ASQ:SE, and was only weakly correlated with the PSI's Difficult Child Scale in the replication sample.

It is notable that the BPSC correlated most strongly with the PSI's Difficult Child Scale, which includes items that directly assess child behavior. The PSI's Dysfunctional Interaction Scale, which mostly includes items that assess the child's effect on the parent, did not correlate with any BPSC subscale. Similarly, the PHQ-2, which is a measure of parent depressive symptoms, correlated with BPSC subscales in the original but not in the replication sample. Because neither the PSI's Dysfunctional Interaction Scale nor the PHQ-2 directly assess constructs targeted by the BPSC, these findings are not unexpected.

Normative Data

DIF analyses suggested a significant effect of age on several items. In addition, analyses in Mplus indicated that the latent factor corresponding to "inflexibility" varied significantly by age. Based on these results, we decided to account for age when calculating normative curves for each subscale. Changes in normative data by age were first examined using Lowess curves. We then fit quantile regression curves to scores of each subscale corresponding to the 50th, 70th, and 90th percentiles. Specifically, age-in-months and a quadratic term were entered into quantile regression equations with each BPSC subscale score as a dependent variable. Results are displayed in Figure 2. No effect of age was identified for "irritability." "Inflexibility" scores

Table 2. Factor Structure

Item	Standardized Factor Loadings*			Thresholds†	
	Irritability	Inflexibility	Difficulty With Routines	Between "Not at All" and "Somewhat"	Between "Somewhat" and "Very Much"
1. Does your child cry a lot?	.817			.653	2.132
2. Does your child have a hard time calming down?	.846			.992	2.629
3. Is your child fussy or irritable?	.823			.847	2.334
4. Is it hard to comfort your child?	.711			1.404	2.293
5. Does your child have a hard time being with new people?		.884		−.140	1.257
6. Does your child have a hard time in new places?		.885		.351	1.835
7. Does your child have a hard time with change?		.835		.535	2.064
8. Does your child mind being held by other people?		.742		−.328	1.162
9. Does your child have trouble staying asleep?			.745	.145	1.478
10. Is it hard to keep your child on a schedule or routine?			.670	.426	1.675
11. Is it hard to put your child to sleep?			.925	.225	1.363
12. Is it hard to get enough sleep because of your child?			.834	.182	1.349

Note: Model estimated in the combined primary care and replication samples; To account for the Baby Pediatric Symptom Checklist (BPSC) 3-option (non-continuous) response scale, we modeled responses to all BPSC items as categorical variables.

*With categorical indicators, factor loadings can be interpreted as in standard factor models.

†With categorical indicators, thresholds are estimated for each item rather than intercepts. The first threshold indicates the score on the latent variable at which the most likely observed response shifts from 0 ("not at all") to 1 ("somewhat"), whereas the second threshold indicates the latent score at which the most likely observed response shifts from 1 (ie, "somewhat") to 2 (ie, "very much").

rose with age before leveling off. "Difficulty with routines" declined over the first months of life before leveling off. In the absence of a "gold standard" criterion to which to compare the BPSC, it is difficult to make firm recommendations regarding appropriate screening thresholds. Whenever possible, we recommend tracking children's scores longitudinally using the charts in Figure 2, thus facilitating a comparison to normative data to determine which children might benefit from further evaluation. As a proxy, a cutoff score of 3 falls between the 70th and 90th percentile for all 3 subscales at all ages and may be considered as an appropriate threshold.

DISCUSSION

Pediatric screening and surveillance for emotional and behavioral disorders is growing. The Individual with Disabilities Education Act of 2004, Bright Futures, and the American Academy of Pediatrics' Task Force on Mental Health emphasize the importance of identifying social-emotional irregularities as early as possible to ensure prompt intervention and recommend behavioral health screening for all children seen in clinical pediatric settings.[47] As of January 1, 2008, all child health care providers in the state of Massachusetts are required to conduct regular behavioral screening for patients on Medicaid at every well-child visit from birth to 21 years of age. This policy was initiated in response to the Rosie D. v Mitt Romney[21] suit filed on behalf of 8 Massachusetts families who felt that the state was failing in its obligation to screen, diagnose, and treat children for mental health disorders. Many pediatricians have instituted the same procedure for patients with private insurance coverage as well.

In order to comply with these recommendations, pediatricians need an instrument that identifies risk for social-emotional disorders among young children that is also brief, easy to administer and to score, and is freely available. Our goal in creating the BPSC was to provide such an instrument. We believe that the BPSC (and its partner, the PPSC[29]) will be feasible for use in primary care settings, and the results presented above provide initial evidence for its use.

The BPSC has many strengths. Its factor structure is well-defined, reliability was adequate across subscales

Table 3. Cronbach's Alpha and Concurrent Pearson Correlations

BPSC Subscale	Original Sample			Replication Sample					
		Pearson Correlation			Pearson Correlation				
							PSI		
	Cronbach's Alpha	ASQ:SE	PHQ-2	Cronbach's Alpha	ASQ:SE	PHQ-2	Parental Distress	Dysfunctional Interaction	Difficult Child
Irritability	.75	0.51**	0.38**	.64	0.24**	0.02	0.22**	0.13	0.42**
Inflexibility	.83	0.09	0.03	.80	0.02	0.15	0.17*	0.10	0.18*
Difficulty with routines	.78	0.42**	0.25**	.74	0.31**	0.04	0.20*	0.07	0.32**
Scale alpha	–	0.70	0.80	–	0.61	0.89	0.85	0.86	0.85

BPSC = Baby Pediatric Symptom Checklist; ASQ:SE = Ages & Stages Questionnaire: Social-Emotional; PHQ-2 = 2 item Patient Health Questionnaire; PSI = Parenting Stress Index; Scale alpha = Cronbach's alpha for each criterion instrument.

Note: *$P < .05$, **$P < .01$.

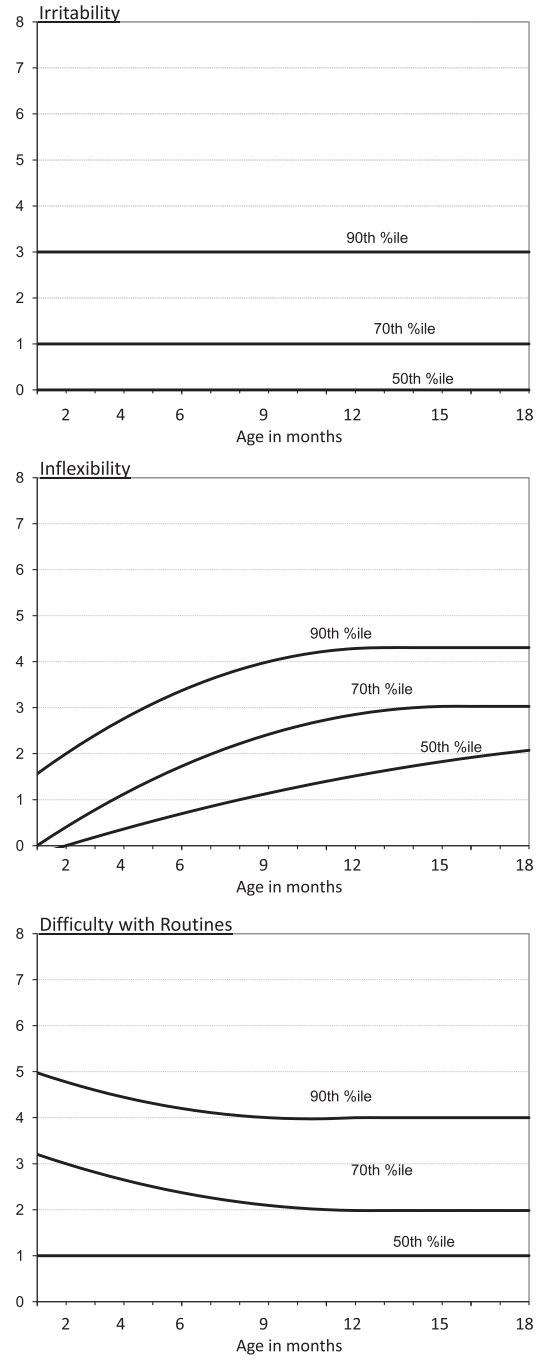

Figure 2. *Normative curves for Baby Pediatric Symptom Checklist (BPSC) scales. Note: Curves are a function of age (months) and age*^{2}*. Constants, linear, and quadratic coefficients for each curve are as follows: irritability 50th = (0, 0, 0); irritability 70th = (1, 0, 0); irritability 90th = (3, 0, 0); inflexibility 50th = (−0.393, 0.204, −0.004); inflexibility 70th = (−0.432, 0.446, −0.014); inflexibility 90th = (1.086, 0.495, −0.019); routines 50th = (1, 0, 0); routines 70th = (3.426, −0.231, 0.009); and routines 90th = (5.2, −0.233, 0.011).*

and across studies, comparing well to the reliability of the ASQ:SE, which is probably the most prominent behavioral screening instrument appropriate for children under 18 months. As their content overlaps with other instruments that assess behavior and temperament, the 3 subscales possess considerable face validity, each assessing a construct that is highly relevant for families with young children. Concurrent correlations suggest that the "irritability" and "difficulty with routines" subscales assess constructs that are similar to those assessed by the ASQ:SE. Despite the "inflexibility" subscale's high internal reliability, low correlations with other instruments suggest that it assesses an area not targeted by the other scales used in this study, and, therefore, that its use requires further study. Finally, normative curves for each subscale by age may facilitate anticipatory guidance, helping parents understand their child's behavior in the context of normative expectations.

Several limitations to this study should be noted. First, our study relies on the comparison of the BPSC to other parent-report instruments (eg, the ASQ:SE and PSI), rather than a "gold standard" criterion assessment, and, therefore, sensitivity and specificity of the BPSC with respect to a known criterion could not be calculated. Unfortunately, although significant advances have been made in the classification of social and emotional problems of very young children (eg, the Zero-to-Three classification system[48]), there is no consensus regarding assessment of behavioral disorders in the first year and a half of life. To further validate the BPSC, future studies could compare BPSC scores between children referred for mental health services and a primary care population. However, identifying such a referred population in this age group may be challenging. To truly validate the BPSC, a prospective longitudinal study of later outcomes is appropriate and necessary and will require a significant commitment of time and funding. Until research with an appropriate criterion can be carried out, the BPSC results can only be interpreted based on available normative data and imperfect knowledge of concurrent validity.

Second, we were not able to enroll all parents who sought pediatric care for their children. Seventy-two percent of eligible parents from waiting room samples and 51% of parents identified from medical records both enrolled and completed study materials. The latter number represents a minimum estimate, because we were unable to contact 32 families—30% of the sample. We may have failed to reach these parents because they did not want to participate, or because their contact information had changed—it is impossible to assess which, or how this may have biased the sample. Moreover, average income and education level was higher in our sample than for the United States as a whole. Although analyses of differential item functioning cannot definitively prove that the BPSC's performance does not vary by gender, race, ethnicity, or parent education, they do suggest that any differences found in the future are not likely to be large.

Despite these limitations, we were able to create a brief scale that is feasible for use with parents of young children and that demonstrates good reliability in comparison with a much longer existing screening instrument. We recommend further validation of the BPSC as a social/emotional screening instrument for infants and toddlers, ideally compared to both concurrent and longitudinal outcomes. In the future, the BPSC would optimally be used as a part of the comprehensive instrument, the SWYC, which

assesses cognitive, motor, and language development as well as behavior and emotional status.

ACKNOWLEDGMENTS

We are indebted to The Commonwealth Fund for its support of the larger project of which this is a part. In particular, Dr Ed Schor's vision and imagination motivated us to create a practical instrument to guide surveillance in pediatric primary care settings. Dr Michael Jellinek had the foresight long ago to create a tool to help pediatricians identify children who had unmet needs for mental health attention. We are grateful also for the assistance of Dr Robin Adair, Dr Liz McGowan, and a large number of pediatricians and other child health specialists,* who allowed us to enroll parents through their offices, to Kate Mattern and Robyn Della Giustina for help with data collection and management, and to the Tufts Medical Center CTSI for ongoing support. We acknowledge as well the wise suggestions of anonymous reviewers of an earlier version of this article.

*__Primary Care Clinics__: Harvard Vanguard Medical Associates; Medical Associates Pediatrics; Revere Health Care Center; Pediatric Healthcare Associates; Porter Pediatrics; Southborough Medical Group; Floating Hospital General Pediatrics; Wilmington Pediatrics; Quality Kids Care; Codman Square Health Center; Westwood-Mansfield Pediatric Associates; and Dr Babu Pediatrics.

Referral Clinics: Floating Hospital Specialty Clinics, including the Center for Children with Special Needs, NICU Follow-Up clinic, Child Psychiatry, and International Adoption Clinic; Occupational Therapy Associates Watertown, South Shore Therapies; Massachusetts General Hospital Child Psychiatry; University of Massachusetts NICU Follow-Up & Developmental-Behavioral Pediatrics Clinics; and Boston Medical Center Developmental Assessment Clinic, Harvard Vanguard Speech/Language Pathology.

SUPPLEMENTARY DATA

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.acap.2012.08.003.

REFERENCES

1. Briggs-Gowan MJ, Carter AS, Skuban EM, et al. Prevalence of social-emotional and behavioral problems in a community sample of 1- and 2-year-old children. *J Am Acad Child Adolesc Psychiatry.* 2001;40:811–819.
2. Kagan J, Snidman N, Zentner M, et al. Infant temperament and anxious symptoms in school age children. *Dev Psychopathol*. 1999;11:209–224.
3. Chronis-Tuscano A, Degnan KA, Pine DS, et al. Stable early maternal report of behavioral inhibition predicts lifetime social anxiety disorder in adolescence. *J Am Acad Child Adolesc Psychiatry.* 2009;48:928–935.
4. Dixon SD, Stein MT. *Encounters with children: pediatric behavior and development*. Philadelphia, Pa: Mosby Inc; 2006.
5. Rubin KH, Burgess KB, Dwyer KM, et al. Predicting preschoolers' externalizing behaviors from toddler temperament, conflict, and maternal negativity. *Dev Psychol*. 2003;39:164–176.
6. Feldman R, Greenbaum CW, Yirmiya N. Mother-infant affect synchrony as an antecedent of the emergence of self-control. *Dev Psychol*. 1999;35:223–231.
7. Degnan KA, Calkins SD, Keane SP, et al. Profiles of disruptive behavior across early childhood: contributions of frustration reactivity, physiological regulation, and maternal behavior. *Child Dev.* 2008;79:1357–1376.
8. Degnan KA, Henderson HA, Fox NA, et al. Predicting social wariness in middle childhood: the moderating roles of childcare history, maternal personality and maternal behavior. *Soc Dev.* 2008;17:471–487.
9. Nyman G. Infant temperament, childhood accidents, and hospitalization. *Clin Pediatr (Phila)*. 1987;26:398–404.
10. Agras WS, Hammer LD, McNicholas F, et al. Risk factors for childhood overweight: a prospective study from birth to 9.5 years. *J Pediatr.* 2004;145:20–25.
11. Spitz AS, Weber-Gasparoni K, Kanellis MJ, et al. Child temperament and risk factors for early childhood caries. *J Dent Child (Chic)*. 2006;73:98–104.
12. Horwitz SH, Gary LA, Briggs-Gowan MJ, et al. Do needs drive services or do services drive needs? *Pediatrics*. 2003;112(6 Pt 1):1373–1378.
13. Leaf PJ, Alegria M, Cohen P, et al. 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.* 1996;35:889–897.
14. Teerikangas OM, Aronen ET, Martin RP, et al. Effects of infant temperament and early intervention on the psychiatric symptoms of adolescents. *J Am Acad Child Adolesc Psychiatry.* 1998;37:1070–1076.
15. Olds DL, Henderson CR Jr, Kitzman H. Does prenatal and infancy nurse home visitation have enduring effects on qualities of parental caregiving and child health at 25 to 50 months of life? *Pediatrics*. 1994;93:89–98.
16. Ammerman RT, Putnam FW, Stevens J, et al. An open trial of in-home CBT for depressed mothers in home visitation. *Matern Child Health J*. 2011;15:1333–1341.
17. Regalado M, Halfon N. Primary care services promoting optimal child development from birth to age 3 years: review of the literature. *Arch Pediatr Adolesc Med*. 2001;155:1311–1322.
18. Meisels SJ, Shonkoff JP. *Handbook of Early Childhood Intervention*. New York, NY: Cambridge University Press; 1990.
19. Drotar D, Stancin T, Dworkin P, et al. Selecting developmental surveillance and screening tools. *Pediatr Rev*. 2008;29:e52–e58.
20. Gardner W, Kelleher KJ, Wasserman R, et al. Primary care treatment of pediatric psychosocial problems: a study from pediatric research in office settings and ambulatory sentinel practice network. *Pediatrics*. 2000;106:E44.
21. Kuhlthau K, Jellinek M, White G, et al. Increases in behavioral health care screening in pediatric care for Massachusetts Medicaid patients. *Arch Pediatr Adolesc Med*. 2011;165:660–664.
22. Jellinek M, Murphy JM, Burns BJ. Brief psychosocial screening in outpatient pediatric practice. *J Pediatr.* 1986;109:371–378.
23. Simonian SJ, Tarnowski KJ. Utility of the Pediatric Symptom Checklist for behavioral screening of disadvantaged children. *Child Psychiatry Hum Dev.* 2001;31:269–278.
24. Ishizaki T, Fukai Y, Kobayashi Y. Utility of the Japanese version of the Pediatric Symptom Checklist to screen children with psychosocial problems in the primary and junior high schools and juvenile guidance and consultation offices in Japan (in Japanese). *Journal of Japanese Society of Psychosomatic Pediatrics*. 2002;10:119–127.
25. Jutte DP, Burgos A, Mendoza F, et al. Use of the Pediatric Symptom Checklist in a low-income, Mexican American population. *Arch Pediatr Adolesc Med*. 2003;157:1169–1176.
26. Gardner W, Lucas A, Kolko DJ, et al. Comparison of the PSC-17 and alternative mental health screens in an at-risk primary care sample. *J Am Acad Child Adolesc Psychiatry.* 2007;46:611–618.
27. Jellinek MS, Murphy JM, Little M, et al. Use of the Pediatric Symptom Checklist to screen for psychosocial problems in pediatric primary care: a national feasibility study. *Arch Pediatr Adolesc Med*. 1999;153:254–260.
28. Jellinek MS, Froehle MD. *Bright futures in practice: mental health*. Arlington, VA: National Center for Education in Maternal and Child Health; 1998.
29. Sheldrick RC, Henson BS, Merchant S, et al. The Preschool Pediatric Symptom Checklist (PPSC): development and initial validation of a new social/emotional screening instrument. *Acad Pediatr.* 2012;12:456–467.
30. Briggs-Gowan MJ, Carter AS, Irwin JR, et al. The Brief Infant-Toddler Social and Emotional Assessment: screening for social-emotional problems and delays in competence. *J Pediatr Psychol*. 2004;29:143–155.
31. Squires J, Bricker D, Twombley E. *The ASQ: SE User's Guide*. Baltimore, MD: Paul H Brookes Publishing Co; 2002.

32. Greenspan S. *Greenspan Social-Emotional Growth Chart*. San Antonio, TX: Harcourt Assessment; 2004.
33. Finello KM, Poulsen MK. The Behavioral Assessment of Baby's Emotional and Social Style (BABES): a new screening tool for clinical use. Presented at the 10th International Conference of Infant Studies; 1996, Providence RI.
34. Neisworth JT, Banato SJ, Salvia J, et al. *TABS Manual for the temperament and atypical behaviors scale: early childhood indicators of developmental dysfunction*. Baltimore, MD: Paul H Brookes Publishing Co; 1999.
35. Baker BL, McIntyre LL, Blacher J, et al. Pre-school children with and without developmental delay: behavior problems and parenting stress over time. *J Intellect Disabil Res*. 2003;47(Pt 4-5):217–230.
36. Donenberg G, Baker BL. The impact of young children with externalizing behaviors on their families. *J Abnorm Child Psychol*. 1993;21: 179–198.
37. Wildman BG, Kizilbash AH, Smucker WD. Physicians' attention to parents' concerns about the psychosocial functioning of their children. *Arch Fam Med*. 1999;8:440–444.
38. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *J Psychosom Res*. 2005;58: 163–171.
39. Abidin RR, Zalaquett CP, Wood RJ. *Parenting stress index: a measure of parent-child system*. Lanham, MD: US Scarecrow Education; 1997.
40. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care*. 2003;41:1284–1292.
41. Patrick CJ, Hicks BM, Nichol PE, et al. A bifactor approach to modeling the structure of the psychopathy checklist-revised. *J Pers Disord*. 2007;21:118–141.
42. Krueger RF, Markon KE, Patrick CJ, et al. Linking antisocial behavior, substance abuse, and personality: an integrative quantitative model of the adult externalizing spectrum. *J Abnorm Psychol*. 2007; 116:645–666.
43. Gibbons RD, Rush AJ, Immekus JC. On the psychometric validity of the domains of the PDSQ: an illustration of the bi-factor item response theory model. *J Psychiatr Res*. 2009;43:401–410.
44. Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual Life Res*. 2007;16(Suppl 1):19–31.
45. Sijtsma K. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*. 2009;74:107–120.
46. Yang FM, Tommet D, Jones RN. Disparities in self-reported geriatric depressive symptoms due to sociodemographic differences: an extension of the bi-factor item response theory model for use in differential item functioning. *J Psychiatr Res*. 2009;43: 1025–1035.
47. American Academy of Pediatrics Task Force on Mental Health. Mental health screening and assessment tools for primary care. *Pediatric*. 2010;125(Supp 3):S173-S192.
48. DC:0-3R. *Diagnostic classification of mental health and developmental disorders of infancy and early childhood*. Revised Edition. Washington, DC: ZERO TO THREE Press; 2005.