



Research report

The stability of parental bonding reports: A 20-year follow-up

Eleanor Murphy*, Priya Wickramaratne, Myrna Weissman

Department of Psychiatry, College of Physicians and Surgeons, Columbia University & Division of Epidemiology, New York State Psychiatric Institute, United States

ARTICLE INFO

Article history:

Received 3 July 2009

Received in revised form 1 January 2010

Accepted 6 January 2010

Available online 6 February 2010

Keywords:

Parenting
Depression
Measurement
Reliability
Family
Memory

ABSTRACT

Background: Addressing the long-term reliability of retrospectively assessed parenting is underscored by the well-documented association between parenting behaviors, and mood disorders in offspring. The rarity of longitudinal research with follow-up periods exceeding 10 years creates a need for additional studies.

Methods: 134 offspring of depressed and non-depressed parents were assessed on Parental Bonding Instrument (PBI) scores, lifetime major depression (MDD), and current depressive symptoms at four waves across 20 years. PBI rank order and mean level stability, individual trajectories, and the impact of baseline age, gender, and lifetime MDD on stability, were obtained using multiple regression and linear mixed model analyses.

Results: Besides paternal overprotection which showed a 1.6-point average decrease, the PBI domains remained non-significant for mean level change over 20 years. However, there was a significant individual variation for all PBI domains. Lifetime MDD and age did not significantly impact retest correlations; older age at baseline was associated with higher average paternal overprotection. Sons had lower retest correlations than daughters, but did not differ from daughters on mean level stability. Current depressive symptoms were associated with PBI scores, but did not impact the effect of lifetime MDD, gender or age on mean level stability and individual trajectories.

Limitations: Small sample sizes and measuring lifetime MDD as present or absent may have restricted our ability to detect effects of MDD history on PBI stability.

Conclusion: The PBI is a robust measure of an important environmental risk for depressive disorders, and can be variably sensitive to sample characteristics, the passage of time and mood fluctuations. However, this sensitivity does not appear to significantly bias the long-term stability of this instrument.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

A commonly assessed correlate of affective disorders in adolescence and adulthood has been childhood perceptions of parenting made possible through the development of standardized measures (Gerlsma et al., 1990; Parker, 1981; Sato et al., 1997). The Parental Bonding Instrument (PBI) is one such measure, utilized to assess offspring's retrospective perceptions of their parents' style of affection and attachment to them (Parker et al., 1979). On this measure, which

primarily evaluates two domains – parental affection and overprotection, individuals are asked to recall how they were parented during their first 16 years of life. Like other retrospectively designed measures, the PBI's long-term stability has been central to many studies, particularly in light of its widely documented association with psychiatric illness, specifically mood disorders.

The robust link between major depressive disorders in offspring, and perceived parental bonding as assessed by the PBI, as been universally demonstrated (Gotlib et al., 1988; Parker et al., 1995; Patton et al., 2001; Plantes et al., 1988; Rey, 1995). In particular, low parental affection has been associated with major depression in offspring (Mackinnon et al., 1993; Parker et al., 1995; Patton et al., 2001). Empirically

* Corresponding author. NYSPI, Unit 24, 1051 Riverside Drive, New York, NY 10032, United States.

E-mail address: em713@columbia.edu (E. Murphy).

supported hypotheses have suggested that the relationship involves multiple pathways, including parental depression which impacts their bonding style and which may also be passed on to offspring (Parker et al., 1992; Rodgers, 1996), and offspring temperament which influences their ability to elicit parental care and also shapes their recollections of parental bonding style (Kendler, 1996; Lichtenstein et al., 2003). Because depression can bias memory towards negative events (Beck, 2002; Dalgleish and Watts, 1990), the relationship between parental bonding styles and depression could potentially be confounded by memory. The mood-congruent memory bias of depression theory supports the idea that individuals tend to recall material that matches the mood they are in during retrieval (Blaney, 1986; Bower, 1981, 1987; Bower and Forgas, 2000; Dalgleish and Watts, 1990). However, with cross-sectional studies in which assessments are done at one point in time, determining the extent to which mood may impact recall remains a challenge.

To address the limitations of cross-sectional designs, longitudinal studies have tracked individuals' mood changes in relation to parental bonding scores over varying periods of time, ranging from a few months (Parker, 1981; Plantes et al., 1988; Gillham et al., 2007) to many years (Lizardi and Klein, 2005; Wilhelm, et al., 2005). Although several studies have demonstrated the PBI's significant consistency and immunity to mood changes over time (Gerlsma et al., 1993; Gerlsma et al., 1994; Lizardi and Klein, 2005, Wilhelm et al., 2005), some have yielded mixed findings (Gillham et al., 2007; Reese et al., 2002), or found significant relationships with PBI and depression change scores. Some limitations of these studies include relatively short follow-up periods (Gillham et al., 2007; Parker, 1981), which could potentially increase the propensity of subjects to recall their earlier responses; female-only samples (Gotlib et al., 1988; Neale et al., 1994), which decreases external validity for male offspring and/or paternal bonding; patient samples (Lizardi and Klein, 2005), which decreases generalizability to non-clinical subjects, and samples that are relatively homogeneous in education and professional occupations (Wilhelm et al., 2005), who may have a better appreciation than less educated subjects, for the need to be consistent in reporting. Further, with the exception of Wilhelm et al.'s 20-year follow-up study (2005), to date there are few, if any published studies tracking stability of PBI reports over extended periods of 10 or more years.

Studies examining PBI stability have traditionally used retest correlations or ANOVA to test mean level stability. However, mean level stability only provides general systematic trajectory of the entire sample, which potentially could mask the effects of individuals changing in opposite directions across time (Vaidya et al., 2008). Retest correlations provides estimates of individual change over time, but does not provide a sense of individual change relative to sample change, nor gives an estimate of change relative to baseline position.

In this study, we utilized a non-clinical sample in order to examine the 20-year stability of the PBI. In addition, we examined the influence of gender, lifetime history of Major Depressive Disorder (MDD), and current depressive symptoms on mean level stability, and individual-level change over time. In addition to using the traditional methods, we

also extend our analyses to include individual growth curves which provide an estimation of individual change relative to the sample, as well as individuals' change overtime relative to their baseline position on the PBI domains (DeLucia and Pitts, 2006; Lenzenweger et al., 2004; Vaidya et al., 2008).

2. Method

2.1. Sample

The sample consisted of 81 females and 53 males who were part of the second generation of a three-generation family study of depression. The original sample included depressed participants who were recruited from a mental health treatment facility in New Haven, CT., as well as "normal" controls selected based on a community survey and based on two systematic interviews, showed no evidence of psychiatric disorder ever. The probands, who were assessed for lifetime depression based on structured clinical interviews, comprised the first generation, which represented 91 families. The probands and their offspring (second and third generations) were followed longitudinally to track a wide variety of physical and mental health outcomes. The original sample and its characteristics are described extensively elsewhere (Weissman et al., 1987, 1997, 2005).

This study focused on the second generation – the only one in which members were assessed at all four waves, across a 20-year period (at baseline, 2, 10 and 20 years later). Additionally, we selected those individuals from that generation who had participated in all four waves. Consequently, the study sample comprised 68% ($n=81/120$) of the original female sample, and 62% ($n=53/85$) of the original male sample. The selected sample represented 69 families, and did not differ significantly from the full sample by PBI or CES-D scores. They did not differ significantly from the full sample by parental depression (OR 0.73, 95% CI = 0.529–1.01), and were slightly more likely to have had a lifetime history of major depression at any time during the period of assessment (OR = 1.33, 95% CI = 0.995–1.77).

The 69 families in our sample contained up to five members, with 42 families having two or more members. The mean number per family was 1.94 subjects. The initial assessment, wave 1, was completed in 1982; wave 2, two years later; wave 3 was completed in 1992 and wave 4 was completed in 2002. The average age at initial assessment was 16.4 years ($SD=4.70$), with females ($M=17.3$ years, $SD=4.33$) significantly older than males ($M=15.0$ years, $SD=4.92$), $t(132)=-2.94$, $p=.004$.

2.2. Measures

2.2.1. Lifetime history of Major Depression (MDD)

Major Depression (MDD) was diagnosed and scored using the semi-structured clinical interview, the Schedule of Affective Disorders and Schizophrenia – Lifetime Version, (SADS-LA, Endicott and Spitzer, 1978). This instrument was developed to increase reliability in psychiatric diagnoses, using symptom criteria that correspond to the Diagnostic and Statistical Manual of Psychiatric Disorders (DSM). The SADS-LA assesses the severity of current psychopathology, as well as the frequency, duration and severity of past psychiatric

illnesses. Diagnoses of MDD made through SADS interviews exhibited high inter-rater reliability, with kappas of .90 (Spitzer, et al., 1978). The SADS-LA is well suited to research in which accuracy of diagnostic ratings is essential.

For this study, masters and doctoral level clinicians, who underwent extensive training and reliability evaluations, administered these interviews. In-depth descriptions of the interview procedures are published elsewhere (Weissman et al., 1987, 1997, 2005). The original SADS-LA was modified between waves 2 and 4 to reflect the ongoing changes in DSM criteria. For children between ages 6 and 17 years, the child-modified version of the SADS-LA was used. Participants were considered to have a lifetime history of MDD, if they were diagnosed with one or more episodes of depression by the last wave of assessment. Sixty-nine (21 males and 48 females) individuals had a lifetime history of major depression. Participants without a history of major depression (non-MDD) were slightly younger ($M = 15.6$ years, $SD = 4.75$) at initial assessment those with a history of major depression (MDD) ($M = 17.2$ years, $SD = 4.56$), $t(132) = -2.00$, $p = .05$. Females were twice as likely to have a lifetime history of major depression as males ($OR = 2.22$, $95\% CI = 1.09-4.49$), which was similar to the gender ratio for MDD history of the original sample ($OR = 2.48$, $95\% CI = 1.40-4.39$).

2.2.2. Current depressive symptoms

The Center for Epidemiological Studies-Depression Scale (CES-D, Radloff, 1977) was used to evaluate current depressive symptoms at each wave. This measure is a self-report 20-item instrument designed to measure the levels of depression symptomatology in the general population. Unlike the SADS-LA, it assesses the extent of a depressed mood along a quantitative dimension, but it is not designed for diagnosing or assessing the severity of Major Depressive Disorder. Each item is scored on a 4-point Likert scale range from “0 = rarely or none of the time (less than 1 day),” to 3 = “most of the time (5–7 days).” An example of an item is “I was bothered by things that usually don't bother me.” Scores could range from 0 to 60, with higher scores indicative of more depressive symptoms. The CES-D is widely used in epidemiological studies on depression, has been tested on a variety of populations (Radloff, 1977), has excellent internal consistency (Radloff, 1977), and adequate discriminant and current validity (Bradburn, 1969; Dohrenwend and Dohrenwend, 1974; Radloff, 1977). The CES-D modified for children (CES-DC) was also used at waves 1 and 2, while the CES-D was used at wave 4. As expected, participants with a history of MDD had higher CES-D scores than those without MDD history (statistically significant at $p < 0.05$, at waves 1 and 4).

2.2.3. Parental bonding

On the Parental Bonding Instrument, participants were asked to recall how they were parented by their mother and father during their first 16 years of life. The 25 items, assessing the domains of affection (12 items) and overprotection (13 items), were scored on a 4-point Likert type scale, ranging from 0 = “very unlike” to 3 = “very like.” For some items, the scale was reversed scored based on the wording of the item. An example of an item scored on the “affection” domain is “could make me feel better when I was upset;” an item scored on the “overprotection” domain is “tried to

control everything I did.” Participants were required to respond to items for mothers and fathers separately. Scores for each domain represent the sum total of domain items and could range from 0–36 in parental affection, and 0–39 on parental overprotection. According to cutoff points established by Parker et al. (1979), scores below 27/24 indicate low maternal/paternal affection, while scores above 13.5/12.5 indicate high maternal/paternal overprotection. The PBI has been evaluated extensively for its psychometric properties, has been used with a variety of populations, and has demonstrated good retest reliability, internal consistency, and validity. (Parker, 1989, 1990).

2.3. Statistical analyses

The data were analyzed using moderated regression techniques and linear mixed models (SAS Proc Mixed), in order to have flexibility in modeling the covariance structure of the random effects of individual change over time. For all analyses where gender was used, we coded males as 0 and females as 1. For lifetime MDD, we coded those with a lifetime history of MDD as 1, and those without MDD history as 0. Time was modeled as number of years, where wave 1 = 0, wave 2 = 2, wave 3 = 10 and wave 4 = 20. For all models that tested the effects of age, we used baseline age instead of age at each wave to avoid confounding the effects of time, since age at each wave is a linear function of baseline age plus the number of years elapsed.

2.4. Rank order stability

Rank order stability was assessed by obtaining the Pearson product-moment correlations between waves. To assess group differences in retest correlations, we performed moderated regression analyses. To test the effect of MDD as a moderator of the PBI correlation between waves, we performed multivariate regression analyses, using PBI scores at preceding waves, MDD history, and the interaction term between MDD history and PBI scores at preceding waves as predictors, and PBI scores at successive waves as the dependent variable. To investigate the impact of gender and baseline age on parental bonding retest correlations, we performed the same analyses as described for MDD history, this time using gender or baseline age as the moderators.

2.5. Individual growth curves

To examine individual trajectories for each PBI domain, we specified an unconditional linear mixed model taking into account the fixed effect of time, and random effects of individual subjects. The first model was analyzed using two levels, looking at each observation per individual (level 1), which were the PBI scores at the four time points assessed, and treating each individual as a separate unit (level 2). We regressed each PBI domain on time as a fixed effect, along with the random effects of individual subjects across time with an unstructured covariance. This model could be specified by the equation, $PBI_{ij} = b_{0i} + b_{1i}(\text{time}_{ij}) + \nu_{0i} + \nu_{1i} + \epsilon_{ij}$, where the PBI score in each domain for the subject “i” at time j was a function of the sample baseline score (b_0), the sample slope (b_1) representing the change in PBI score per unit change in time,

the individual variation from the baseline (ν_0), individual variation from sample slope (ν_1) and error residuals (ϵ_{ij}).

We further explored the effects of using 3 levels using family as a clustering variable. Clarke (2008) used simulated models to examine the effects of clustering on the random and fixed estimates, and standard errors, with unbalanced groups where $n \leq 2$ cases per group. It was shown that for the disaggregated linear models, where clustering was ignored, the fixed effects were little impacted, even though they could potentially be overestimated because of a small downward bias in the standard errors. With the clustered linear models, the fixed effects were not appreciably biased despite having wider confidence intervals, but the group level random effects were more likely to be undetected owing to inflated standard errors leading to imprecise estimates and reduced power. Because our data were relatively sparse and unbalanced, with a mean of 1.9 family members per family, we took into account the possible effects of these constraints in comparing the two models and interpreting the results.

2.6. Mean level stability

To examine mean level stability and the influence of gender, diagnosis, and age on mean level stability, we specified five models: In all models, we controlled for CES-D scores because they were related to all PBI domains at each wave, with higher depressive symptoms associated with lower maternal and paternal affection, and higher maternal and paternal overprotection. The first model examined the stability of each PBI domain across the four waves by assessing the fixed effect of time (in years). This is similar to the unconditional model described above. The second model examined the influence of gender in mean level stability. In this model, we specified gender and the interaction of time and gender as predictors, with CES-D scores as a covariate. In the third model, we examined the influence of diagnosis (lifetime MDD), following the same steps as we did for gender. The fourth model examined the influence of age at entry into the study (baseline age). Finally in the fifth model, we examined the effects of each predictor of interest holding the others constant by entering all predictors simultaneously, controlling for CES-D scores.

3. Results

3.1. PBI rank order stability

As detailed in Table 1, the PBI retest correlations were all significant across the waves. However, they were generally

Table 1
Rank order stability for PBI reports.

PBI domain	Stability coefficient waves ^a					
	W1-2	W1-3	W1-4	W2-3	W2-4	W3-4
Maternal affection	.68	.53	.35	.60	.40	.54
Maternal overprotection	.68	.43	.35	.47	.34	.52
Paternal affection	.59	.45	.33	.62	.38	.55
Paternal overprotection	.53	.38	.23	.56	.35	.56

Note: The correlation coefficients are all statistically significant at $p \leq .05$.

^a W1 = baseline; W2 = 2 years; W3 = 10 years; and W4 = 20 years.

modest with the strength of decreasing in magnitude as the periods between assessments increased. This pattern ensued across all four parental bonding domains.

3.1.1. Effects of gender, lifetime MDD, and age on rank order stability

Table 2 shows the regression coefficients for the moderator effect of MDD history and gender on the parental bonding correlations. Gender appeared to have modest effects in the consistency in reporting of parental bonding. In all instances, correlations for females were stronger in magnitude, and significantly so in about one-third of the assessments.

MDD history showed no significant impact on the retest correlations except for one significant regression coefficient for paternal overprotection (W1–W3), where those with a history of MDD had a higher retest correlation between those waves. In addition, there were no discernable patterns in terms of the magnitude of the correlations, or in terms of the specific parental bonding domains, based on MDD history.

Despite the differences in age at the initial assessment, we found no significant relationships between baseline age and the rank order retest stability of the reports. The regression coefficients (not shown in table) ranged from 0.01 to 0.06 and were not statistically significant.

3.2. PBI mean level stability

Table 3 shows the means and standard errors for the PBI domains at each point of assessment. The results showed little systematic change in mean levels for maternal affection, maternal overprotection and paternal overprotection. In these three domains, the change in average scores from one wave to the next did not exceed 1.6 points, nor was there consistency in the direction of the mean level changes across the waves. While there was a statistically significant decline in mean paternal affection at year 20, the decrease of less than 2 points given an average score of 24.11 across the four time points is relatively small.

3.2.1. Individual variation over a 20-year period

Table 4 details the fixed and random effects for the 2-level unconditional growth model. The fixed intercepts represent the baseline scores for each PBI domain. The slopes represent the average change in scores per year. As we would expect, the fixed effect estimates shown in Table 4 correspond to the results in Table 3. For instance, the intercepts for each domain represent the sample scores at year 0 (baseline), which should approximate the scores in Table 3's baseline column. Furthermore, with the exception of paternal affection, the slope estimates are not statistically significant. The statistically significant slope for paternal affection shows a decrease of only 0.08 points per unit increase in time, less than two points over a 20 year period. When we included CES-D scores as a covariate, the patterns of the results were similar in direction, and statistical significance for all the estimates, which did not change more than ± 0.01 points in magnitude (not shown in table).

The variance of slope estimates are all significant, suggesting that there is significant variation in individual trajectories across the 20-year period. Nonetheless, the effect

Table 2
Impact of gender and lifetime MDD on rank-order stability of PBI reports.

PBI domains	Variables	Unstandardized regression coefficients					
		W1-2 ^a	W1-3	W1-4	W2-3	W2-4	W3-4
Maternal affection	MDD history	−0.08	0.05	−0.03	0.22	−0.12	−0.19
	Gender	0.18	0.30*	0.04	0.35*	0.12	0.27
Maternal overprotection	MDD history	−0.01	−0.06	−0.13	−0.06	−0.07	−0.03
	Gender	0.38**	0.41**	0.15	0.13	0.05	0.24
Paternal affection	MDD history	−0.03	−0.06	−0.04	−0.15	0.05	−0.05
	Gender	0.05	0.06	0.06	0.14	0.17	0.30*
Paternal overprotection	MDD history	0.02	0.32*	0.08	0.26	0.22	−0.05
	Gender	0.28	0.36*	0.14	0.29*	0.10	0.29*

N = 134.

Note. For MDD history, depressed group has higher PBI retest correlations than non-depressed group if regression coefficients are positive, but lower PBI retest correlations if coefficients are negative. For gender, daughters have higher PBI retest correlations than sons if coefficients are positive but lower correlations if coefficients are negative.

^a Wave 1 = baseline; wave 2 = 2 years; wave 3 = 10 years; and wave 4 = 20 years.

* $p \leq .05$.

** $p \leq .01$.

sizes are relatively small. For all PBI domains, there is more individual variation relative to the sample. For maternal affection, an offspring whose change rate is one standard deviation above average is expected to increase at the rate of 0.30 points per year, resulting in an expected total increase of 6 points over the 20 year period. Conversely, for an offspring whose change rate is one standard deviation below average, the maternal affection score is expected to decrease by 0.18 points a year, resulting in an expected total of about 4 points over a 20 year period. For maternal overprotection, there is a 0.33 point increase per year in offspring whose change rate was one standard deviation above average, and a 0.17 point decrease in offspring whose change rate was one standard deviation below average. Variation in change rates of paternal affection and overprotection were of similar magnitude to variation in maternal scores.

The covariances of the intercept and slope are all significant and negative. This indicates that initial position at baseline (year 0) is a significant predictor of the type of change that will occur, and that those individuals who started off with lower scores tend to increase overtime, while those who started off with higher scores tend to decrease over time.

When we compared the results from the 2-level model to that of the 3-level model (clustering of family), we obtained similar fixed effects for both models, but the random effects for individuals (albeit significant) were smaller in the 3-level model. The group (family) level effects were not statistically significant for maternal affection and overprotection, but the slope variance was statistically significant for paternal

affection and marginally significant for paternal overprotection. Because the group level effects may have been underestimated, we caution against interpreting our group level random effects results, given the sparseness of our data.

3.2.2. Effects of gender, lifetime MDD, and age on mean level stability

Table 5 shows the fixed effects of gender, lifetime MDD and age on PBI reports. The results shown are from the conditional linear models 2, 3 and 4 as described in the Method. Results from the full model is not shown in the table, but described below.

Gender had a marginally significant effect on baseline reports of maternal affection and maternal overprotection even after controlling for CES-D scores, with daughters reporting higher maternal affection and lower maternal overprotection than sons. However, gender had no significant influence on the stability of these PBI scores over 20 years, as measured by the interaction between slope and gender in the fixed effects. Gender had no statistically significant effect on paternal affection or overprotection either. In addition there was no variation by gender in PBI scores at baseline for these domains. When gender was examined at part of a fuller model with all the predictors included, the results were similar with a marginally significant effect on baseline reports of maternal affection, $F(130.60) = 3.90$, $p = .05$, maternal overprotection, $F(130.63) = 3.18$, $p = .08$, no significant effects on baseline reports of paternal affection, $F(127.28) = 1.13$, $p = .29$, or paternal overprotection, $F(128.18) = 0.31$,

Table 3
Mean level stability for PBI reports.

Domain	Means				F
	Baseline	Year 2	Year 10	Year 20	
Maternal affection	28.26 (0.57)	28.93 (0.52)	28.27 (0.53)	29.39 (0.52)	2.41
Maternal overprotection	11.37 (0.65)	10.99 (0.65)	12.27 (0.61)	11.33 (0.53)	1.66
Paternal affection	24.34 ^a (0.63)	24.84 ^a (0.62)	24.35 ^a (0.66)	22.92 ^b (0.60)	2.93*
Paternal overprotection	11.75 (0.67)	10.41 (0.63)	10.62 (0.63)	10.14 (0.56)	1.99

N = 134.

Note. Standard errors are in parentheses. For paternal affection, means sharing a superscript are not significantly different at $p < .05$.

* $p < .05$.

Table 4

Growth curve estimates for PBI reports.

Effect	Maternal affection	Maternal overprotection	Paternal affection	Paternal overprotection
<i>Fixed effect estimates</i>				
Intercept	28.41 (0.52)	11.36 (0.61)	24.77 (0.59)	11.16 (0.59)
t-Statistic	54.88***	18.75***	42.25***	19.02***
Slope	0.04 (0.03)	0.02 (0.04)	-0.08 (0.04)	-0.06 (0.04)
t-Statistic	1.26	0.49	-2.38*	-1.54
<i>Random effects estimates</i>				
Variance of intercept	28.24 (4.43)	37.84 (6.10)	34.62 (5.70)	33.90 (5.76)
z-Statistic	6.36***	6.20***	6.08***	5.89**
Variance of slope	0.06 (0.02)	0.07 (0.02)	0.07 (0.02)	0.07 (0.02)
z-Statistic	3.82***	3.27***	3.17***	3.19**
Covariance of intercept and slope	-0.74 (0.22)	-1.14 (0.31)	-0.73 (0.28)	-0.91 (0.30)
z-Statistic	-3.39***	-3.72***	-2.62**	-3.04**
Residual variance	14.36	21.22	21.50	23.00
-2LL	3267.6	3428.1	3451.6	3465.2

N = 134.

Note. Estimates shown are from 2-level unconditional linear mixed model which does not control for CES-D scores. Controlling for CES-D scores did not alter the patterns, direction and statistical significance of the estimates.

* $p \leq .05$.** $p \leq .01$.*** $p \leq .001$.

$p = .58$. Likewise, there were no significant effects for mean level stability on maternal affection, $F(126.07) = 0.18$, $p = .84$, maternal overprotection, $F(127.76) = 0.36$, $p = .70$, paternal affection, $F(123.37) = 0.59$, $p = .56$, and paternal overprotection, $F(126.63) = 0.63$, $p = .53$.

Lifetime MDD was not significantly associated either with baseline PBI scores or stability of PBI scores except for paternal affection. Lifetime MDD was significantly associated with baseline PBI scores for paternal affection and was marginally associated with stability of paternal affection, with 14.5% of the variation in the stability of PBI scores for paternal affection explained by lifetime MDD status. Offspring with lifetime MDD had lower baseline scores for paternal affection, which on average tended to decrease across time, compared to those without lifetime MDD. When lifetime MDD was entered into the full model, the results were similar for maternal affection and overprotection and paternal affection, with no significant effects for baseline PBI scores

or stability for maternal affection and overprotection, and significant effect on baseline scores, and marginally significant effects on stability for paternal affection. However, MDD became significant for baseline scores for paternal overprotection, $F(131.86) = 5.41$, $p = .02$, with the lifetime MDD group showing lower average baseline scores.

Age at entry into study did not have significant effects on the baseline reports or stability of PBI scores for maternal affection and maternal overprotection. Age at entry into study had a marginally significant effect on baseline reports of PBI scores of paternal affection, but not on stability of paternal affection. While there was no effect on baseline reports of paternal overprotection, age at entry was significantly associated with the stability of paternal overprotection, explaining 30% of the variation in the stability. When age at entry was entered in the full model, the effects on baseline scores and mean level stability for each PBI domain were similar to that of the results shown in Table 5.

Table 5

Impact of gender, MDD history, and age on mean level stability of PBI reports.

Fixed effects	Maternal affection	Maternal overprotection	Paternal affection	Paternal overprotection
<i>Gender</i>				
Baseline reports	1.83 (1.05) ⁺	-2.30 (1.25) ⁺	-0.38 (1.13)	0.31 (1.20)
Stability across time	-0.067 (0.06)	0.089 (0.07)	0.040 (0.07)	0.072 (0.07)
<i>MDD history</i>				
Baseline reports	-0.14 (1.04)	0.04 (1.24)	-2.17 (1.10)*	1.75 (1.17)
Stability across time	-0.072 (0.06)	0.038 (0.07)	-0.121 (0.06) ⁺	0.063 (0.07)
<i>Age</i>				
Baseline reports	-0.08 (0.10)	-0.11 (0.12)	-0.19 (0.11)	-0.17 (0.11) ⁺
Stability across time	-0.001 (0.006)	0.010 (0.01)	-0.003 (0.01)	0.020 (0.01)**

Note. Figures represent slope estimates and standard errors are in parentheses. Baseline reports represent slopes for gender, MDD history or age, while stability across time represents slopes for interaction between gender and time, MDD history and time, and age and time. All effects listed are examined separately controlling for CES-D Scores. CES-D scores (not shown) are significantly associated with PBI domains in all models.

⁺ $p \leq .10$.* $p \leq .05$.** $p \leq .01$.

4. Discussion

This study examined the 20-year consistency in reporting of a retrospective measure of parental bonding, and the extent to which gender, depression history and age affected the consistency in reporting. We used three indices of reliability – rank order stability or retest correlations, mean level stability, and individual growth curves. Our findings indicate that the PBI is a reliable assessment tool for measuring an important aspect of the environmental risk of some psychiatric disorders. While there may be significant individual variation in long-term trajectories as evidenced by significant random subject effects and modest retest correlations, these fluctuations are relatively limited in range and tend to converge overtime, both of which have the overall impact of creating significant mean level stability in the overall sample across a 20-year period. Secondly, gender, lifetime MDD and age do not seem to significantly impact long-term mean level stability, although gender has a somewhat stronger influence on rank order stability. In addition, current depressive symptoms are strongly related to the retrospective reports of parental bonding, but do not explain the relationships between time and PBI scores, or the impact of gender, lifetime MDD history and baseline age on PBI stability. This finding suggests that there are other factors not assessed in our study, but which simultaneously exert their influence on current mood and perceived parental bonding in offspring.

Our results showed negligible effects of lifetime MDD on rank order and mean level stability. When we did find significant differences, for instance for paternal affection and overprotection, the differences were small such that with an approximate average baseline score of 24, a change of less than 2 points over a 20-year period is not likely to be clinically meaningful. These findings are consistent with Wilhelm et al.'s (2005) 20-year study, which found no significant differences in PBI retest reliability based on lifetime MDD history.

In our study, a main limitation was that lifetime MDD was dichotomously scored as present (having one or more episodes), versus absent (having no episodes), and therefore did not differentiate between individuals according to the number of depressive episodes. Secondly, the study did not take into account when subjects developed depression – overlooking the possibility that some assessments may have been taken before a given subject developed major depression. However, if the latter limitation was significantly biasing the results, we might expect to see increasingly pronounced MDD group differences with the passage of time. With the exception of a slightly lower score at year 20 for paternal affection, there is no evidence of this trend for the other PBI domains.

In general, daughters' reports showed stronger retest correlations in parental bonding scores compared to that of sons. However, we found no gender differences in mean level stability. These mixed findings are consistent with previous research on retrospective accounts of childhood experiences and perceptions of family dynamics, which has found either no male–female differences in rank order or mean level stability, or some gender differences, where females are more reliable (Widom and Morris, 1997; Wilhelm et al., 2000).

Females may be more reliable in some instances, as research has shown that women are more attuned with their emotional state and feelings (Wilhelm and Parker, 1994), and may be better able at separating perceptions of parenting from current mood states and other influences. Compared to men, women are also more likely to be in personal and professional roles involving childcare, which may serve to maintain the salience of parent–child relationships and therefore reinforce their interest and motivation in responding to family-related inventories.

Although we found no significant influence of age when we assessed reliability from one wave to the next, when we compared the magnitude of the coefficients from early wave pairs to later ones, we found that the correlations between earlier assessments showed a trend of being smaller than those between latter assessments. For example, as observed from Table 3, the strength of the correlations between waves 3 and 4 was greater than that between waves 1 and 3, even though the same number of years elapsed between both sets of waves. Our moderator analyses indicated that older age is associated with stronger correlations, but this association was not statistically significant. The age effects may be real but quite small, and unable to be detected without a large sample. On the other hand, the discrepancy between correlations at W1–3 and W3–4 may not be a function of age per se but instead may be related to a practice effect of repeated assessments. It also may have reflected qualitative changes within families as children transition into adulthood and parents become more settled in their respective roles.

In considering the applicability of our findings, we advise our readers to note the primary limitations of this study, which are the small non-probability sample, and the lumping of individuals with one or more depressive episodes into one group. Although our sample sizes fell within the range of those of recently published studies on PBI reliability (Gillham, et al., 2007; Lizardi and Klein, 2005; Wilhelm et al., 2005), we may not have had sufficient power to detect the effects of MDD history or age on PBI retest reliability. However, our findings are consistent with previous research which found no significant effects on PBI reliability based on MDD history.

Conclusion

The PBI has shown basic reliability with individual fluctuations that may not be clinically meaningful over a 20 year period. The finding of PBI–current mood association does not necessarily point to mood biasing the reports, but may suggest the incorporation of additional events and interpersonal processes that may give rise to the changes in mood and simultaneously influence the interpretation of parents' behavior. Like any instrument that taps human subjectivity across the lifespan, the PBI will reproduce the variation in perception that exists within a given individual over time. As succinctly summarized in Hardt and Rutter's (2004) review on the validity of retrospective measures when considering the possible effects of mood and other influences on recall

“Memory needs to be conceptualized in terms, not of a fixed storehouse of deposited bits of information, but rather of a perception of the past that is open to

influences from attempts to provide meaning, and from cognitive processes involving election, abstraction, interpretation, integration and reconstruction, as well as from the effects of repeated recall and rehearsal (or lack of both).” (Hardt and Rutter, 2004, pp. 268).

Consistent with this view, researchers and practitioners alike are well advised to exercise appropriate judgment, particularly when considering the temporal proximity of reports to events in question, and the specific characteristics of the study or treatment subjects. In treatment-related psychological interventions that commonly solicit perceptions of childhood events, knowing how contemporaneous events affect these memories or are influenced by them can provide valuable insights to both professionals and their patients. Similarly, for clinical studies in which prospective designs are not viable, researchers can benefit by reinforcing retrospective reports through repeated assessments, external measures, and other observations of the operationalized construct that can be characterized independently.

Role of funding source

This study was made possible from grants from the National Institute of Mental Health (NIMH) R01MH036197-19, 5-T32 MH13043-34, and R01MH060912-S06. The NIMH had no further role in the study design; collection analysis and interpretation of data, and writing of the manuscript.

Conflict of interest

M. Weissman – In the past year, Dr. Weissman received research funding from the National Institute of Mental Health (NIMH), the National Institute on Drug Abuse (NIDA), the National Alliance for Research on Schizophrenia and Depression (NARSAD), and the Sackler Foundation; and royalties from the Oxford University Press, Perseus Press, the American Psychiatric Association Press, and MultiHealth Systems. She has no conflict of interest to disclose with this manuscript.

All other authors have no conflict of interest to declare.

Acknowledgements

The authors would like to thank Virginia Warner for her valuable assistance in the provision of the necessary data for this study.

References

Beck, A., 2002. Cognitive models of depression. In: Dowd, T., Leahy, R. (Eds.), *Clinical Advances in Cognitive Psychotherapy: Theory and Application*. Springer Publishing Co, New York, NY, US, pp. 29–61. xiv, 446 pp.

Blaney, P.H., 1986. Affect and memory: a review. *Psychol. Bull.* 99, 229–246.

Bower, G.H., 1981. Mood and memory. *Am. Psychol.* 36, 129–148.

Bower, G.H., 1987. Commentary on mood and memory. *Behav. Res. Ther.* 6, 443–455.

Bower, G.H., Forgas, J.P., 2000. Affect, memory, and social cognition. In: Eich, E., Kihlstrom, J.F., Bower, G.H., Forgas, J.P., Neidenthal, P.M. (Eds.), *Cognition and Emotion*. Oxford University Press, New York, NY, pp. 87–168.

Bradburn, N.M., 1969. The structure of psychological well-being. Aldine, Chicago.

Clarke, P., 2008. When can group level clustering be ignored? Multilevel models versus single-level models with sparse data. *J. Epidemiol. Community Health* 62, 752–758.

Dalgleish, T., Watts, F.N., 1990. Biases of attention and memory in disorders of anxiety and depression. *Clin. Psychol. Rev.* 10, 589–604.

DeLucia, C., Pitts, S.C., 2006. Applications of individual growth curve modeling for pediatric psychology research. *J. Pediatr. Psychol.* 31 (10), 1002–1023.

Dohrenwend, B.S., Dohrenwend, B.P. (Eds.), 1974. *Stressful Life Events: Their Nature and Effects*. Wiley-Interscience.

Endicott, J., Spitzer, R.L., 1978. A diagnostic interview: the schedule for affective disorders and schizophrenia. *Arch. Gen. Psychiatry* 35, 837–844.

Gerlsma, C., Emmelkamp, P.M.G., Arrindell, W.A., 1990. Anxiety, depression, and perception of early parenting: a meta-analysis. *Clin. Psychol. Rev.* 10, 251–277.

Gerlsma, C., Das, J., Emmelkamp, P.M., 1993. Depressed patients' parental representations: stability across changes in depressed mood and specificity across diagnoses. *J. Affect Disord.* 27, 173–181.

Gerlsma, C., Kramer, J.J., Scholing, A., Emmelkamp, P.M., 1994. The influence of mood on memories of parental rearing practices. *Br. J. Clin. Psychol.* 33, 159–172.

Gillham, J.E., Putter, P., Kash, V.M., 2007. The effects of sad mood on reports of parents' caregiving. *J. Psychopath. Behav. Assess.* 29, 125–133.

Gotlib, I.H., Mount, J.H., Cordy, N.I., Whiffen, V.E., 1988. Depression and perceptions of early parenting: a longitudinal investigation. *Br. J. Psychiatry* 152, 24–27.

Hardt, J., Rutter, M., 2004. Validity of adult retrospective reports of adverse childhood experiences: review of the evidence. *J. Child. Psychol. Psychiatry* 45 (2), 260–273.

Kendler, K., 1996. Parenting: a genetic-epidemiologic perspective. *Am. J. Psychiatry* 153 (1), 11–20.

Lenzenweger, M.F., Johnson, M., Willett, J.B., 2004. The longitudinal study of personality disorders. *Arch. Gen. Psychiatry* 61, 1015–1024.

Lichtenstein, P., Ganiban, J., Neiderhiser, J.M., Pedersen, N.L., Hansson, K., Cederblad, M., Elthammar, O., Reiss, D., 2003. Remembered parental bonding in adult twins: genetic and environmental influences. *Behav. Genet.* 33 (4), 397–408.

Lizardi, H., Klein, D., 2005. Long-term stability of parental representations in depressed outpatients utilizing the parental bonding instrument. *J. Nerv. Ment. Dis.* 193 (3), 183–188.

Mackinnon, A., Henderson, A.S., Andrews, G., 1993. Parental 'affectionless control' as an antecedent to adult depression: a risk factor refined. *Psychol. Med.* 23, 135–141.

Neale, M.C., Waters, E., Heath, A.C., Kessler, R.C., Perusse, D., Eaves, L.J., Kendler, K.S., 1994. Depression and parental bonding: cause, consequence, or genetic covariance? *Genet. Epidemiol.* 11, 503–522.

Parker, G., 1981. Parental reports of depressives: an investigation of several explanations. *J. Affect. Disord.* 3, 131–140.

Parker, G., 1989. The Parental bonding instrument: psychometric properties reviewed. *Psychiatr. Dev.* 7 (4), 317–335.

Parker, G., 1990. The parental bonding instrument: a decade of research. *Soc. Psychiatry Psychiatr. Epidemiol.* 25, 281–282.

Parker, G., Tupling, H., Brown, L., 1979. A parental bonding instrument. *Br. J. Med. Psychol.* 52, 1–10.

Parker, G., Barret, E.A., Hickie, I.B., 1992. From nurture to network: examining links between perceptions of parenting received in childhood and social bonds in adulthood. *Am. J. Psychiatry* 149 (7), 877–885.

Parker, G., Hadzi-Pavlovic, D., Greenwald, S., Weissman, M., 1995. Low parental care as a risk factor to lifetime depression in a community sample. *J. Affect. Disord.* 33, 173–180.

Patton, G.C., Coffery, C., Posterino, M., Carlin, J.B., Wolfe, R., 2001. Parental 'affectionless control' in adolescent depressive disorders. *Soc. Psychiatry Psychiatr. Epidemiol.* 36, 475–480.

Plantings, M.M., Prusoff, B.A., Brennan, J., Parker, G., 1988. Parental representations of depressed outpatients from a U.S.A. sample. *J. Affect. Disord.* 15 (2), 149–155.

Radloff, L.S., 1977. The CES-D scale: a self-report depression scale for use in the general population. *Appl. Psychol. Meas.* 1, 385–401.

Reese, R.J., Kieffer, K.M., Briggs, B.K., 2002. A reliability generalization study of select measures of adult attachment style. *Educ. Psychol. Meas.* 62, 619–646.

Rey, J.M., 1995. Perceptions of poor maternal care associated with adolescent depression. *J. Affect. Disord.* 34 (2), 95–100.

Rodgers, B., 1996. Reported parental behaviour and adult affective symptoms: two mediating factors. *Psychol. Med.* 26, 63–77.

Sato, T., Sakado, K., Uehara, T., Nishioka, K., Kasahara, V., 1997. Perceived parental styles in patients with depressive disorders: replication outside Western culture. *Br. J. Psychiatry* 170, 173–175.

Spitzer, R.L., Endicott, J., Robins, E., 1978. Research diagnostic criteria for a selected group of functional disorders. New York: Biometric Research Unit. New York Psychiatric Institute.

Vaidya, J.G., Gray, E.K., Haig, J.R., Mroczek, D.K., Watson, D., 2008. Differential stability and individual growth trajectories of big five and affective traits during young adulthood. *J. Pers.* 76 (2), 267–304.

Weissman, M., Gammon, G., John, K., Merikangas, K., Warner, V., Prusoff, B., Sholomskas, D., 1987. Children of depressed parents: increased psychopathology and early onset of major depression. *Arch. Gen. Psychiatry* 44, 487–533.

Weissman, M., Warner, V., Wickramaratne, P., Moreau, D., Olfson, M., 1997. Offspring of depressed parents: ten years later. *Arch. Gen. Psychiatry* 54, 932–940.

Weissman, M., Wickramaratne, P., Nomura, Y., Warner, V., Verdelli, H., Pilowsky, D., Grillon, C., Bruder, G., 2005. Families at high and low risk of

- depression: a three-generation study. *Arch. Gen. Psychiatry* 62 (1), 29–36.
- Widom, C.S., Morris, S., 1997. Accuracy of adult recollections of childhood victimization: part 2. childhood sexual abuse. *Psychol. Assess.* 9 (1), 34–46.
- Wilhelm, K., Parker, G., 1994. Sex differences in depression rates: fact or artefact? *Psychol. Med.* 24, 97–111.
- Wilhelm, K., Brownhill, S., Boyce, P., 2000. Marital and family functioning: different measures and viewpoints. *Soc. Psychiatry Psychiatr. Epidemiol.* 35, 358–365.
- Wilhelm, K., Niven, H., Parker, G., Hadzi-Pavlovic, D., 2005. The stability of the parental bonding instrument over a 20-year period. *Psychol. Med.* 35, 387–393.