A Prospective Test of the Stress-Buffering Model of Depression in Adolescent Girls: No Support Once Again

Emily Burton and Eric Stice
University of Texas at Austin

John R. Seeley
Oregon Research Institute

The stress-buffering model posits that social support mitigates the relation between negative life events and onset of depression, but prospective studies have provided little support for this assertion. The authors sought to provide a more sensitive test of this model by addressing certain methodological and statistical limitations of past studies with prospective data from 496 adolescent girls. Deficits in peer support predicted increases in depressive symptoms, and negative life events predicted onset of depressive pathology. However, none of the 14 prospective tests provided support for the stress-buffering model despite sufficient power. Results provide scant support for the stress-buffering model and suggest that it might be time to shift attention to alternative multivariate models concerning these risk factors.

Major depression is one of the most common psychiatric conditions to afflict adolescents and is characterized by a recurrent course and elevated comorbid psychopathology (Lewinsohn, Hops, Roberts, Seeley, & Andrews, 1993; Newman et al., 1996; Stice, Presnell, & Bearman, 2001). Diagnostic and subdiagnostic depression increase the risk for future onset of psychiatric disorders, substance abuse, academic failure, suicide attempts, legal problems, and social impairment (Gotlib, Lewinsohn, & Seeley, 1995). Research has focused on identifying factors that increase the risk for the onset of depression in an effort to (a) elucidate the etiologic processes that give rise to this pernicious disorder, (b) inform the content of prevention interventions, and (c) identify risk factors for selecting high-risk populations for targeted prevention programs.

One of the most widely studied psychosocial risk factors for depression is life stress (i.e., negative life events). Individuals who experience a single negative life event (e.g., parental divorce or relationship breakup) may have little difficulty garnering appropriate coping resources (i.e., money to cope with financial crisis, psychological resources to cope with loss, etc.). Such coping resources enable persons to more effectively deal with stressors or more quickly recover from them (Wheaton, 1983; Zeidner & Hammer, 1990). However, when an individual experiences multiple negative life events, their coping resources are stressed (e.g., relationship breakup and serious health problem). Multiple stressors may overburden coping resources, putatively putting that individual at risk for the development of psychological disorders including depression (S. Cohen & Wills, 1985; Kessler, 1979; Zeidner & Hammer, 1990).

Negative life events have been found to prospectively predict increases in depressive symptoms in children, adolescents, and adults1 (L. H. Cohen, Burt, & Bjorck, 1987; Compass, Howell, Phares, Williams, & Giunta, 1989; DuBois, Felner, Brand, Adan, & Evans, 1992; Lewinsohn et al., 1994; Monroe, Imhoff, Wise, & Harris, 1983; Nolen-Hoeksema, Girmus, & Seligman, 1992; Windle, 1992). However, null findings have been observed in studies with smaller samples (Monroe, 1983; Siegel & Brown, 1988; Swearingen & Cohen, 1985; Zimmerman, Ramirez-Valles, Zapert, & Maton, 2000). An important gap in the literature is that the majority of studies have focused on predicting change in continuous symptom measures rather than onset of clinically significant depressive pathology. We located only two studies with adolescents that examined onset of depressive pathology (Goodyer, Herbert, Tamplin, & Altham, 2000; Lewinsohn, Allen, Seeley, & Gotlib, 1999); in both studies major life stress predicted onset of major depression. A benefit of examining onset of depression is that it ensures that the risk factors are predicting clinically meaningful levels of pathology. Accordingly, one aim of this study was to test whether negative life events predicted onset of depressive pathology.

It is also noteworthy that the effect sizes for negative life events are relatively small—explaining an average of only 2% of the variance in subsequent increases in depressive symptoms (e.g., Garber, Kelley, & Martin, 2002; Lewinsohn et al., 1994; Nolen-Hoeksema et al., 1992). This suggests that most of the variance in

---

1 It is important to note that we include only prospective studies that control for initial levels of the outcome variables because only these clearly establish temporal precedence (i.e., model change).
depression is accounted for by other factors. Previous theorists have posited that differences in the magnitude of negative life events may explain the inconsistent findings. However, research suggests that the decomposition of negative life events into subcategories of major events and daily hassles resulted in no differences in the prediction of increases in depressive symptoms (Robinson, Garber, & Hilsenrath, 1995; Windle & Windle, 1996). Because it appears that the predictive effects for major events and daily hassles are generally comparable, and because excluding trivial negative life events is thought to increase the validity of negative life event measures (Clear, 1981), we focused solely on major negative life events in this study.

Another widely studied psychosocial risk factor for depression is social support. It has been argued that individuals who possess greater social resources (i.e., supportive friends and family members) and who share a high degree of intimacy with those resources (i.e., confidants) are less likely to experience depression (S. Cohen & Wills, 1985; Pierce, Frone, Russell, Cooper, & Mudar, 2000). This is because if an individual is accepted and valued in their social environment, they may be more likely to feel greater esteem, confidence, and efficacy—factors which could guard against the development of depression (Nezlek, Kowalski, Leary, Blevins, & Holgate, 1997).

Deficits in social support have been found to prospectively predict increases in depressive symptoms in adolescents (DuBois et al., 1992; Lewinsohn et al., 1994; Sheeber, Hops, Alpert, Davis, & Andrews, 1997; Windle, 1992; Zimmerman et al., 2000), but nonsignificant findings have been reported in adult samples (Monroe, 1983; Monroe et al., 1983). We find it interesting that this effect is more consistently observed among women than among men (Lewinsohn, Hoferman, & Rosenbaum, 1988; Slavin & Ranier, 1990; Windle, 1992). This latter finding might occur because women are at greater risk for the development of depression (Ge, Lorenz, Conger, Elder, & Simons, 1994). Additionally, Windle (1992) found that, by separating out parental support from peer support, only parental support was a significant predictor of increases in depressive symptoms. Therefore, we examined the predictive power of these two facets of support separately. However, like the effects for negative life events, the effect sizes for social support are relatively small, accounting for an average of about 2% of the variance in increases in depressive symptoms (e.g., Slavin & Ranier, 1990; Windle, 1992). Also, it is important to note within this context that it is always possible that some omitted third variable explains prospective findings such as those between low social support and subsequent increases in depressive symptoms. For example, this relation might have occurred solely because some other risk factor for depression, such as high-trait neuroticism, increases the risk for both low social support and onset of depression.

The stress-buffering hypothesis proposes that social support attenuates the relation between negative life events and the risk for development of depression (S. Cohen & Wills, 1985; Wheaton, 1985). This interactive model posits that, when faced with troubling life events, individuals with greater support from family and friends are less likely to become depressed than individuals with lower levels of support. This social support presumably enhances efficacy, esteem, and confidence, thereby increasing an individual’s perception that he or she can cope effectively with negative life events. In addition, the tangible support provided by network members may directly facilitate the resolution of negative life events (e.g., financial assistance).

There is a general consensus in both research and clinical circles that the stress-buffering model has been empirically supported (e.g., Brown & Harris, 1978; S. Cohen & Wills, 1985; Leavy, 1983). We find it somewhat surprising, however, that prospective studies have provided little support for the stress-buffering effect in the prediction of increases in depressive symptoms. In studies with adults, only 1 of the 36 stress-buffering interactions tested by Monroe et al. (1983) was significant, only one of the two stress-buffering interactions tested by McFarlane, Norman, Streiner, and Roy (1983) was significant, and none of the eight stress-buffering effects tested by Monroe (1983) was significant. In studies with adolescents, only 1 of the 10 stress-buffering interactions tested by DuBois et al. (1992) was significant and neither of the two tests for this hypothesis by Zimmerman et al. (2000) was significant.

Given that only 3 of the 58 stress-buffering interactions tested in these prospective studies were significant, which is expected on the basis of chance alone (all studies used a .05 alpha level), it is possible that these findings are simply Type I errors. However, alternative explanations for the modest support for the stress-buffering model should be considered. First, the studies might have had insufficient power, in that the average sample size across studies was only 202 participants (range = 75–428). This conjecture is consistent with the fact that the study with the largest sample size generated the most support for the stress-buffering model (McFarlane et al., 1983). Power might have also been limited because of measurement unreliability, as the reliability of the cross-product interaction term used to test an interaction is a product of the reliability of the constituent main effect terms (Aiken & West, 1991). It was difficult to assess the veracity of this explanation because most researchers did not report information about the reliability of the measures used in their studies.

A second possible explanation for the fact that virtually all past tests of the stress-buffering model produced null findings is that the length of follow-up used in these studies may not have been optimal for capturing the stress-buffering effect. If length of follow-up was related to the likelihood of finding support for the stress-buffering model, then studies that used particularly short (or particularly long) follow-up periods should have generated positive findings (i.e., length of follow-up should have been correlated with the likelihood of finding effects). This was not the case, as both studies that used very brief follow-up periods (e.g., 1 month) and those that used very long follow-up periods (e.g., 2 years) did not provide support for this model.

A third explanation for the consistent lack of support for the stress-buffering model is that there were specific problems with the measures used to assess the key variables. The finding that structured-interview measures of negative life events are more valid than self-report surveys (McQuaid, Monroe, Roberts, Kupfer, & Frank, 2000) suggests that studies that used the former type of measure might be more likely to find support for this model. However, both studies that used structured-interview and self-report survey measures of negative life events generated null findings. It is also possible that more consistent support for the stress-buffering model might have emerged if certain aspects of social support networks were assessed (e.g., network density vs. perceived support). However, null findings have been observed for studies that used a wide variety of support indices. Thus, it would
appear that the null findings cannot be easily ascribed to the limitations of a particular measurement instrument.

A final possible explanation for the weak support of the stress-buffering model is that all of the studies predicted change in continuous measures of depressive symptoms rather than onset of clinically significant depressive pathology. Focusing on predicting onset of clinically significant depressive pathology may provide a more powerful test of this hypothesis (the greater homogeneity across participants in the criterion reference group should reduce error variance and facilitate separation from participants with more transient depressive symptoms that are not of syndromal form).

Accordingly, in an effort to provide a more sensitive test of the stress-buffering model, we tested for this interaction in a large sample and used reliable measures of support and life stress. In addition, we tested whether negative life events and deficits in social support evidence main and interactive effects in the prediction of both increases in continuous measures of depressive symptoms and onset of depressive pathology. We used diagnostic interviews in an effort to ensure that we were assessing only clinically significant depressive pathology. We also tested for stress-buffering effects using classification tree analysis (Breiman, Friedman, Olshen, & Stone, 1984; Kiernan, King, Kraemer, Stefanick, & Killen, 1998) because this statistical technique is more sensitive to detecting interactions than standard approaches. We examined parental and peer support separately because of the evidence that effects may differ for these two sources of support (Windle, 1992). We thought it optimal to test these relations in the context of a young adolescent sample because depressive symptoms and negative life events increase markedly during this developmental period (Cole et al., 2002; Ge et al., 1994; Newcomb, Huba, & Bentler, 1981). We focused exclusively on female adolescents because these data were drawn from a longitudinal study of the risk factors for eating pathology, which predominantly affects adolescent girls. Nonetheless, we felt this was an appropriate sample because adolescent girls show higher rates of depression and negative life events than adolescent boys (Ge, Conger, & Elder, 2001; Ge et al., 1994).

Method

Participants

We recruited 496 adolescent girls from public and private middle schools in a metropolitan area of the southwestern United States. Adolescents ranged in age from 11–15 (M = 13.0, SD = 0.73; mode = 13) at baseline. The sample was composed of 2% Asians-Pacific Islanders; 7% African Americans; 68% Caucasians; 18% Latinas; 1% Native Americans; and 4% who specified other or mixed racial heritage, which was representative of the ethnic composition of the schools from which we sampled (2% Asians-Pacific Islanders; 8% African Americans; 65% Caucasians; 21% Latinas-Latinos; 4% other or mixed). Parental education of the participants, a proxy for socioeconomic status, was as follows: 29% high school graduate or less; 23% some college; 33% college graduate; and 15% graduate degree, which was representative of the educational attainment of comparably aged adults from the census (34% high school graduate or less; 25% some college; 26% college graduate; 15% graduate degree). Thus, the sample included a wide range of participants who were primarily from lower to middle class backgrounds.

Procedures

The study was described to parents and participants as an investigation of adolescent mental and physical health. An active parental consent procedure was used, wherein an informed consent letter and a stamped self-addressed return envelope were sent to all parents of girls in the seventh and eighth grades in the participating schools (a second mailing was sent to nonresponders after 2 weeks). Adolescent assent was also secured immediately before data collection took place. This resulted in an average participation rate of 56% of eligible students across schools. This participation rate was similar to that of other school-recruited samples that used active consent procedures and involved structured interviews (e.g., 61% for Lewinsohn et al., 1993). The 1-year prevalence rates of major depression (4.2%), bulimia nervosa (0.4%) and substance abuse (8.9%; Stice, Presnell, & Bearman, 2001) were also similar to the prevalence rates from other epidemiological studies (Lewinsohn et al., 1993; Newman et al., 1996).

Girls completed a questionnaire and participated in a structured interview at baseline (T1) and three annual follow-up assessments (T2, T3, and T4). However, because negative life event data were not collected at T1, analyses focused on data from T2, T3, and T4. Female assessors with a bachelor’s, master’s, or doctorate degree in psychology conducted interviews. Assessors attended 24 hrs of training in which they (a) were taught structured interview skills, (b) reviewed diagnostic criteria for relevant Diagnostic and Statistical Manual of Mental Disorders (4th ed.; DSM-IV; American Psychiatric Association, 1994) disorders, (c) observed simulated interviews, and (d) role-played interviews. Assessors had to demonstrate an interrater agreement (κ > .80) with experts using tape-recorded interviews before collecting data. A randomly selected subset of interviews (5% of those conducted annually) were audiotaped so that ongoing supervision could be provided regarding interview techniques. Assessments took place during regular school hours or immediately after school on the school campus or at participants’ houses. Girls received a $15 gift certificate to a local book and music store to compensate them for participating in the study.

Measures

Perceived social support. Perceived social support was measured with 12 items from the Network of Relationships Inventory (Furman & Buhrmester, 1985). Items assessed companionship, guidance, intimacy, affection, admiration, and reliable alliance from parents and peers. Research has provided evidence that this measure possesses internal consistency (α = .89), test–retest reliability (r = .69), and predictive validity (Furman, 1996; Furman & Buhrmester, 1985; Stice & Bearman, 2001). Pilot testing (n = 30) revealed a 1-week test–retest correlation of .95 for the perceived parental support scale and .90 for the perceived peer support scale. The internal consistency was .90 for the perceived parent support scale and .88 for the perceived peer support scale at T2.

Negative life events. The Major Life Events scale (Lewinsohn et al., 1994) was used to assess the occurrence of nine negative life events during the past year. This nine-item scale has been found to evidence acceptable predictive validity (Lewinsohn et al., 1994). Researchers have long recognized (Cleary, 1981) that internal consistency is not an appropriate index of the reliability for stressful life events measures because experiencing one negative event (e.g., having a possession stolen) should not increase the odds of experiencing others (e.g., experiencing an illness). Consistent with this assertion, the Cronbach’s alpha for our sample (.51) was similar to those for other widely used stressful life events measures (alphas ranged from .40–.53; Hurst, Jenkins, & Rose, 1978). Pilot testing (n = 40) revealed a 1-week test–retest correlation of .90.

Depressive symptoms. The Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS; Puig-Antich & Chambers, 1983), a structured psychiatric interview, assessed the diagnostic symptoms of DSM-IV major depression. Our version of the K-SADS was
based on that used by Lewinsohn et al. (1993), which combined features of the epidemiological and the present episode versions. Responses were used to classify participants as having met threshold or subthreshold diagnostic criteria for major depression during the past year at T2 (4.2% and 5.4%, respectively), T3 (1.6% and 4.8%, respectively), and T4 (5.1% and 5.7%, respectively). Girls who reported the presence of at least five of the symptoms necessary for a diagnosis, but who endorsed a subthreshold level on at least one of these symptoms, were given a subthreshold diagnosis (DSM-IV depressive disorder not otherwise specified). The interviewer had to rate the symptoms as clinically significant for a threshold and subthreshold diagnosis. Items were also standardized (to control for the different response formats of the items) and summed to form overall T2, T3, and T4 depression symptom composites. The test–retest reliability (κ = .73–1.00), internal consistency (α = .68–.85), and discriminant validity of this measure have been documented in past studies (Ambrosini, 2000; Lewinsohn et al., 1994; Puig-Antich & Chambers, 1983). A randomly selected subset of participants (5%) were interviewed within a 3-day period by a second clinical assessor who was blind to the first diagnosis, resulting in high interrater agreement (κ = 1.00). Another randomly selected subset of participants (5%) completed a second diagnostic interview with the same clinical assessor 1 week later, resulting in excellent test–retest reliability (κ = 1.00). The continuous depression scale (sum of depression ratings by each participant) evidenced acceptable internal consistency at T2 (α = .83).

Analytic Overview

Preliminary analyses tested for attrition biases and examined the correlations between predictor and criterion variables. We then tested whether main effects of perceived support and negative life events and their interaction predicted increases in the continuous measure of depressive symptoms. This was accomplished with growth curve models that tested whether the T2 independent variables predicted subsequent growth in the depressive symptoms over the 2-year follow-up period (i.e., depressive symptom slope), controlling for initial levels of depressive symptoms (i.e., depressive symptom intercept). A unique feature of this analytic technique is that it fits an individual growth curve for each participant with all available data, which provides an optimally reliable method of modeling change in continuous variables (Bryk, Raudenbush, Cheong, & Congdon, 2000). Next, we tested whether T2 main effects of perceived support and negative life events and their interaction predicted onset of depressive pathology (subthreshold or full-threshold major depression) over the 2-year follow-up period among participants initially free of depressive pathology. Finally, we conducted ancillary analyses to address the possibility that the stress-buffering interactions might have emerged if alternative analytic techniques were used or different measures examined.

Predicting Change in a Continuous Measure of Depressive Symptoms

Random regression growth curve models tested for main and interactive effects for T2 perceived support and negative life events in the prediction of growth in depressive symptoms over the 2-year follow-up period, controlling for initial depressive symptoms. The slope coefficients represent the average linear change per year in the continuous depressive symptom measure for each participant. The depressive symptom intercept parameter was coded to represent the value of the linear growth trajectory for each participant at T2. In these models the T2 main effect for perceived support and negative life events and the depressive symptom intercept parameter were entered on the first step and the Support × Negative Life Events interaction term was entered on the second step. The main effect variables were mean centered before

Table 1

<table>
<thead>
<tr>
<th>Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>M</th>
<th>SD</th>
<th>Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. T2 negative life events</td>
<td>-.23*</td>
<td>-.01</td>
<td>.35*</td>
<td>.12*</td>
<td>.17*</td>
<td>1.24</td>
<td>1.30</td>
<td>1.29</td>
<td></td>
</tr>
<tr>
<td>2. T2 parental support</td>
<td>-.16*</td>
<td>-.32*</td>
<td>.10*</td>
<td>.11*</td>
<td>3.90</td>
<td>0.92</td>
<td>-0.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. T2 peer support</td>
<td>-.27*</td>
<td>.04</td>
<td>-.07</td>
<td>4.35</td>
<td>0.75</td>
<td>-1.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Depressive symptom intercept</td>
<td>-.50*</td>
<td>.33*</td>
<td>1.35</td>
<td>0.37</td>
<td>1.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Depressive symptom slope</td>
<td>-.41*</td>
<td>0.03</td>
<td>0.20</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Onset of depressive pathology</td>
<td>-.13</td>
<td>0.33</td>
<td>2.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. T2 = second annual follow-up assessment.
*p < .05.
being used to form the cross-product terms used to test for the interactions for these and all subsequent analyses (Aiken & West, 1991). Separate models were estimated for parental support and peer support. The unstandardized regression coefficients, 95% confidence intervals, percentage of variance explained, and significance levels are reported in Table 2.

Perceived peer social support, but not parental social support, showed a significant unique relation to subsequent increases in depressive symptoms. Negative life events did not show a significant unique relation to future increases in depressive symptoms (although this effect was nonsignificant in the peer support model). Neither the Parental Support \times Negative Life Events nor the Peer Support \times Negative Life Events interaction was statistically significant. The significant effect for peer support was small in magnitude.

**Predicting Onset of Depressive Pathology**

Hierarchical logistic regression models tested whether negative life events and social support showed significant main and interactive effects in the prediction of depressive pathology onset. These models predicted onset of subthreshold and threshold depression over a 2-year follow-up period among adolescents who did not meet criteria for subthreshold or threshold depression at T2. Of the 435 initially nondisordered girls at T2, 57 showed onset of major depression or subthreshold depression over the 2-year follow-up period (33 cases were subthreshold depression and 24 cases were threshold depression). Paralleling the above models, the main effects for negative life events and social support were entered on the first step and the interaction term was entered on the second step. Separate models were estimated for parent and peer support. The unstandardized regression coefficients, 95% confidence intervals, odds ratios, percentage of variance explained, and significance levels are reported in Table 3.

Negative life events increased the risk for onset of depressive pathology, but perceived parental and peer support did not show significant unique relations to subsequent onset of depressive pathology. The Parental Support \times Negative Life Events and the Peer Support \times Negative Life Events interactions did not reach significance. The significant effect for negative life events was small in magnitude.\(^2\)

**Ancillary Analyses**

We conducted a series of ancillary analyses to address the possibility that the stress-buffering interactions might have reached statistical significance if different analytic techniques were used or different measures examined. First, we estimated auto-regressive models that tested for main and interactive effects for T2 perceived support and negative life events in the prediction of T3 depressive symptoms, controlling for T2 depressive symptoms. Hierarchical multiple regression models were used for this purpose, wherein T2 main effects for perceived support and negative life events and T2 depressive symptom scores were entered on the first step and the Support \times Negative Life Events interaction was not statistically significant in the growth curve or onset analyses when the 15 participants with threshold or subthreshold anorexia nervosa or bulimia nervosa at T2 were excluded.

---

2 It should be noted that the rates of threshold or subthreshold anorexia nervosa or bulimia nervosa were 15 (3.1%) at T2, 18 (3.8%) at T3, and 16 (3.2%) at T4 (see Stice, Presnell, & Bearman, 2001 for details regarding diagnostic criteria). None of the stress-buffering interactions were statistically significant in the growth curve or onset analyses when the 15 participants with threshold or subthreshold anorexia nervosa or bulimia nervosa at T2 were excluded.
tion term was entered on the second step. Although peer social support and negative life events showed significant relations to future increases in depressive symptoms over the 1-year follow-up, neither the Parental Support × Negative Life Events nor the Peer Support × Negative Life Events interaction was significant.

Second, we conducted change score analyses that predicted change in the continuous measure of depression from T2–T3 (i.e., T3 depressive symptoms – T2 depressive symptoms) to test whether this analytic technique might produce support for the stress-buffering interaction. Paralleling the results from the other analyses, neither the Parental Support × Negative Life Events interaction nor the Peer Support × Negative Life Events interaction was significant.

Third, to rule out the possibility that the absence of stress-buffering interactive effects resulted because we did not combine perceived support from parents and peers, we reestimated the prospective models using such a composite support variable. Paralleling the results from the other analyses, neither the Parental Support × Negative Life Events interaction nor the Peer Support × Negative Life Events interaction was significant.

Finally, we conducted classification tree analyses (CTA; Breiman et al., 1984; Kierman et al., 1998) to test for evidence of a stress-buffering interaction because this analytic technique is more sensitive to detecting interactions than the conventional analyses used above. CTA uses an empirically based recursive partitioning approach that selects the optimal cutpoint on the optimal risk factor (of all possible cutpoints and risk factors) for generating subgroups with differential risk for a dichotomous outcome. This procedure is then repeated in each successive subgroup until there are no remaining risk factors that identify subgroups at significantly differential risk or until the node sizes become too small. When different risk factors emerge for two branches from the same fork (i.e., the optimal predictor for the outcome in one subgroup is different than that for another subgroup), this signifies an interaction. Separate models tested for interactions between parental support and negative life stress and between peer support and negative life stress in the prediction of onset of depressive pathology among initially nondepressed adolescents. To minimize capitalization on chance, we set the minimum terminal node size to be 20 or greater (alpha was set at .05).

The parental support model produced a tree with one fork and two terminal nodes. Negative life events was the only variable to identify subgroups at significant differential risk for onset of depressive pathology, $\chi^2(1, N = 431) = 11.73, p < .001$. Participants who reported 2.5 or more negative life events showed a higher probability for experiencing onset of depressive pathology ($p = .26$) than participants who reported fewer negative life events ($p = .10$). No significant predictors of depressive pathology onset emerged for either of the first two branches. The peer support model produced an identical tree with only one fork and two terminal nodes. Negative life events was again the only variable to identify subgroups at significantly differential risk for onset of depressive pathology (same split value and $\chi^2$ value). Because there were no statistically significant differential splits on the two forks of these models, there was no evidence of an interaction between perceived parental support and negative life events or between perceived peer support and negative life events in the prediction of depressive pathology onset. That is, even in the more sensitive analyses there was no evidence for the stress-buffering model.

Discussion

Replicating earlier findings (e.g., Lewinsohn et al., 1999), our results showed that adolescents reporting elevated negative life events showed significantly greater risk for onset of depressive pathology. Although some past studies have not observed this relation, these studies had relatively less statistical power because

Table 3

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$B$</th>
<th>95% CI for $B$</th>
<th>Odds ratio</th>
<th>% variance explained</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parental support models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 parental support</td>
<td>-.25</td>
<td>-.56, .07</td>
<td>0.78</td>
<td>.00</td>
<td>.119</td>
</tr>
<tr>
<td>T2 negative life events</td>
<td>.33</td>
<td>.11, .55</td>
<td>1.39</td>
<td>.02</td>
<td>.004</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 parental support</td>
<td>-.18</td>
<td>-.52, .16</td>
<td>0.84</td>
<td>.00</td>
<td>.304</td>
</tr>
<tr>
<td>T2 negative life events</td>
<td>.29</td>
<td>.05, .53</td>
<td>1.34</td>
<td>.01</td>
<td>.016</td>
</tr>
<tr>
<td>T2 Parental Support × Negative Life Events</td>
<td>-.13</td>
<td>-.35, .09</td>
<td>0.88</td>
<td>.00</td>
<td>.240</td>
</tr>
<tr>
<td><strong>Peer support models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 peer support</td>
<td>-.31</td>
<td>-.69, .07</td>
<td>0.73</td>
<td>.00</td>
<td>.112</td>
</tr>
<tr>
<td>T2 negative life events</td>
<td>.38</td>
<td>.16, .60</td>
<td>1.46</td>
<td>.03</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 peer support</td>
<td>-.30</td>
<td>-.70, .10</td>
<td>0.74</td>
<td>.00</td>
<td>.143</td>
</tr>
<tr>
<td>T2 negative life events</td>
<td>.38</td>
<td>.16, .60</td>
<td>1.46</td>
<td>.03</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>T2 Peer Support × Negative Life Events</td>
<td>-.13</td>
<td>-.43, .17</td>
<td>0.88</td>
<td>.00</td>
<td>.394</td>
</tr>
</tbody>
</table>

Note. T2 = second annual follow-up assessment; CI = confidence interval.
of their smaller sample sizes. Thus, our results add support to the theory that negative life events increase the risk for the development of depression. That the effect for negative life events emerged in the models predicting onset of depressive pathology, but not in those predicting change in depressive symptoms, provides support for our suggestion that analyses predicting onset of clinically significant pathology can be more powerful than models predicting change in continuous symptom measures.

We found evidence that peer support, but not parental support, predicted increases in depressive symptoms. It was noteworthy that Windle (1992) found effects for parental, but not peer, support. Although it is possible that these differential findings solely arose from chance, it might also be the case that differences in the samples and measures explain the contradictory findings. In contrast to our sample, Windle’s sample contained men (39%), was primarily Caucasian (97%), and was approximately 2 years older at baseline. Windle also used different measures of social support and depression (e.g., we used a structured-diagnostic interview and he used a self-report measure). It would be useful if future research investigated which sources of support show the strongest relations to depressive pathology and identified factors that qualify these relations (e.g., gender).

Not 1 of the 14 separate tests of the stress-buffering interaction was significant in the current study. These findings converge with the fact that 55 of the past 56 prospective tests of the stress-buffering model with adolescent and adult samples produced null findings. Because 69 of the 72 prospective tests of this model have produced null findings, this suggests that this model may be invalid. However, it is important to carefully consider alternative explanations for the nonsignificant findings because it is hazardous to retain the null hypothesis. First, it is possible that studies have lacked the power to detect interaction effects. The reliability of the interaction term is the product of the reliabilities of the measures used to assess the two main effect variables and therefore can be quite low (Aiken & West, 1991). Because most of the past studies did not report the reliabilities of the constituent measures, it is difficult to assess whether they had sufficient power to detect interaction effects. Nonetheless, the 1-week test–retest reliability coefficients for the measures used in the current study were .95, .90, and .90 for parental support, peer support, and negative life events, respectively. Therefore, the Parental Support \times Negative Life Events interaction term reliability was .86 and the Peer Support \times Negative Life Events interaction term reliability was .81. Even after attenuating for the unreliability of the interaction term and the base rate for onset of depressive pathology, our power to detect a small effect (2% variance explained) for the interaction term was greater than .80. Thus, it is unlikely that insufficient power explained the absence of support for the stress-buffering interaction in the current study.

A second possible explanation for the null findings is that the 2-year follow-up period might have been too long to detect the processes articulated by the stress-buffering model. Stronger support for this interactive model might have emerged if a shorter follow-up period had been used. However, post hoc analyses indicated that this model was not supported in models that used a 1-year follow-up period. In addition, past tests of the stress-buffering model that used follow-up periods ranging from 1 month to 6 months (McFarlane et al., 1983; Monroe, 1983; Monroe et al., 1983; Zimmerman et al., 2000) also found nonsignificant stress-buffering interactions. Collectively, these results provide little support for the assertion that this model would have been supported if a shorter follow-up period had been examined.

Third, it is possible that the null findings for the stress-buffering model resulted because we used a self-report measure of negative life events with questionable validity. Research has found that there is only moderate concordance between self-report measures of negative life events and those that are coded as negative life events in structured interviews (McQuaid et al., 1992, 2000). Thus, the imprecision introduced by idiosyncratic appraisals of the significance and impact of life events might have obscured our ability to detect stress-buffering effects. Nevertheless, the fact that past studies that used structured interviews to assess negative life events (e.g., McFarlane et al., 1983; Monroe et al., 1983) have also generated nonsignificant stress-buffering interactions seems to suggest this is an unlikely explanation for the null findings. In addition, despite the fact that there is evidence that excluding daily hassles from inventories of negative life events increases the validity of life events measures (Cleary, 1981), it is possible that our exclusive focus on major events impacted our findings. However, other researchers that included life events inventories with varying degrees of length and complexity, in both interview and self-report form, also generated null findings (e.g., Monroe et al., 1983; Zimmerman et al., 2000). Again, it seems that this factor alone cannot explain the lack of support for the model.

Fourth, stronger support for the stress-buffering model might have emerged if we had used a measure of social support that reflected different aspects of support (e.g., enacted support or network density), rather than one that reflected perceived support. However, the fact that past studies that assessed various aspects of social support have generated null effects for the stress-buffering model (McFarlane et al., 1983; Monroe, 1983; Monroe et al., 1983) also suggests this explanation cannot easily account for the null findings. More generally, it should also be noted that the fact that we observed significant main effects for our measures of negative life events and social support suggests that the null stress-buffering findings cannot be easily ascribed to limitations of these measures (e.g., a restriction in range).

The current findings also have an important methodological implication. The fact that past cross-sectional studies have provided support for the stress-buffering model (e.g., Brown, Bhrolchain, & Harris, 1975; S. Cohen & Hoberman, 1983; Paykel, Emmes, Fletch, & Rassby, 1980), whereas 96% of the prospective tests of this model have produced null findings, illustrates the hazards of relying on cross-sectional data. With cross-sectional data there is simply no way of determining whether an independent variable is a precursor, consequence, or concomitant of the dependent variable, effectively rendering cross-sectional findings uninterpretable.

Although this study attempted to improve on earlier research by using a prospective design, structured diagnostic interviews, and a large sample, several limitations should be noted. First, we relied heavily on self-report measures. More confidence could have been placed in the findings if we had used structured interviews to assess all constructs and had collected data from multiple informants. Second, the relatively low recruitment rate (56%) for this study suggests that some caution should be used in generalizing these findings. Finally, our use of an entirely female sample limits our ability to generalize these findings to men.
With regard to direction for future research, our opinion is that future studies should conduct more exploratory tests of multivariate models that might explain how various risk factors for depression work in concert. Although it will be necessary to conduct confirmatory tests of any models suggested by exploratory analyses, we suspect that we could learn much from such analyses. It is possible that extant theories may blind us to alternative conceptualizations of how risk factors operate conjointly to produce complex psychiatric conditions such as depression.

In sum, results from our study provided support for the assertions that negative life events and deficits in social support increase the risk for development of depressive pathology, but they also suggest that only certain sources of support had predictive power. More importantly, despite the fact that the stress-buffering hypothesis is widely accepted (e.g., S. Cohen & Wills, 1985; Leavy, 1983), our review of the literature suggested that there was very little prospective support for this interactive model. Although we attempted to provide a more sensitive test of this model by improving on certain limitations of past studies, we still did not find support for the stress-buffering model. This state of affairs suggests that it might be prudent to acknowledge that this intuitively attractive model does not accord with extant prospective findings and implies that we should refocus our research efforts on new etiologic accounts concerning how risk factors may work together to promote depressive pathology.

References


Accepted November 8, 2003