The Independence of Physical Attractiveness and Symptoms of Depression in a Female Twin Population

RENE J. MCGOVERN
Departments of Psychiatry, Neurology, and Psychology
Case Western Reserve University

MICHAEL C. NEALE
KENNETH S. KENDLER
Departments of Psychiatry and Human Genetics
Medical College of Virginia, Virginia Commonwealth University

ABSTRACT. The relationship between physical attractiveness and symptoms of depression was investigated in a general population sample of 1,100 female twins. Photographs were rated by 4 raters. Symptoms of depression were measured by the Depression subscale of the SCL-54, by a self-rating based on the DSM-III-R, and by an MD diagnosis based on a structured interview (SCID). No relationships between ratings of physical attractiveness and symptoms of depression were found.

THE AMOUNT OF MONEY AND TIME spent in the pursuit of beauty in our society suggests that physical attractiveness is not a trivial matter. From a young age, women know that it is important to be beautiful. Even infants and young children show preferences for physical attractiveness in both peers and adults (Langlois et al., 1987; Zakin, 1983). Studies have shown intracultural (Adams, 1977; Cash & Smith, 1982) and cross-cultural (Cunningham, 1986; Langlois et al., 1987; Maret, 1983) reliability of ratings of physical attractiveness, with the attractiveness of the subject having an influence on the attitudes, attributions, and actions of social observers throughout the life span (Adams, 1977; Langlois et al., 1987). Attractiveness has been shown to produce cognitive, affective, and behavioral reactions that are consistent with the “what is beautiful is good” stereotype (Dion, Berscheid, & Walster, 1972).

Research indicates that different expectations occur as a function of appearance (Berscheid & Walster, 1974; Hatfield & Sprecher, 1986; Langlois & Stephan, 1981). Physical attractiveness has been found to be stronger than ethnicity in predicting attributes (e.g., prejudice; Stephan & Langlois, 1984).
Moreover, this stereotype about beauty has been shown to create expectations about mental health. Psychologists, psychiatrists, social workers, and people in general attribute greater psychological disturbance to homely people than to attractive ones (Barocas & Vance, 1974; Cavior & Dokecki, 1971; Farina et al., 1977; Hatfield & Sprecher, 1986; Martin, Friendmeyer, & Moore, 1977; Novotny, 1977). Less attractive people have been shown to expect unattractive people to have poorer mental health (O'Grady, 1982). If unattractive people as well as laypersons and mental health professionals have such expectations, the question remains whether less attractive people actually are more likely to experience symptoms of emotional distress (Bull & Rumsey, 1988).

Reviews of the literature on physical attractiveness (Adams, 1977; Bull & Rumsey, 1988; Cash, 1981; Hatfield & Sprecher, 1986; Patzer, 1985) have documented its importance. Furthermore, researchers have considered the importance of attractiveness on social influence and development (Chaiken, 1986; Sorell & Nowak, 1981). To date, however, there is only one study concerning physical attractiveness in a twin sample (Rowe, Clapp, & Wallis, 1987). Despite various attempts to demonstrate causality between physical attractiveness and symptoms of emotional distress (Farina, Austad, Burns, Bugglin, & Fischer, 1986; Napoleon, Chassin, & Young, 1980), problems with these studies, such as small sample sizes and nongeneralizability of the samples, make interpretation of the results difficult.

In the present study, we examined the relationship between physical attractiveness and symptoms of depression in a large, population-based, female-twin sample. The sample for this study was found through public records rather than through requests for twin volunteers, where self-selection biases exist. The advantages of using such a sample are twofold. First, the results should be more generalizable than results from previous studies using either clinical or college-age participants. Second, the sample was population based and should be more heterogeneous in age, socioeconomic status, and level of education. Thus, measures of attractiveness and symptoms of depression are more likely to be free of bias and other effects of nonrandom sampling.

This study was supported by Grant MH-40828 from the National Institute of Mental Health, Bethesda, MD. The Virginia Twin Registry, established and maintained by W. Nance and L. Corey, is supported by Grants HD-26746 and NS-25630 from the National Institute of Health, Bethesda, MD.

Leroy Thacker assisted in the data analysis. P. Winter and J. Spence assisted in the zygosity diagnoses. Data were collected under the direction of Patsy Waring, assisted by Lou Hopkins. Interviews were conducted by J. Barwick, M. J. Beaman, B. Berry, C. Brankley, C. Browder, E. Carpenter, E. Fay, S. Hall, K. Hough, J. Johnson, V. Johnson, M. Kim, R. Lemon, M. Peery, H. Smith, W. Siddith, E. Tillet, and J. Young. Thanks to Elisabeth Koss, John Bolger, and Carol Prescott for helpful comments on earlier drafts of this article.

Address correspondence to Rene J. McGovern, Department of Psychiatry, Case Western Reserve University, 11100 Euclid Ave., Cleveland, OH 44106-5000.
In addition, the physical attractiveness construct can be further validated by comparing the ratings of monozygotic twins and dizygotic twins. Monozygotic twins have identical genetic makeup (MZ = 1), and dizygotic twins share approximately ½ of their genes (DZ = ½). If, as seems plausible, physical attractiveness is entirely attributable to additive genetic factors, then the correlation should be approximately twice as great in MZ as in DZ twins.

**Method**

**Participants**

This study was part of the larger Virginia Twin Study (VTS) on Stress and Coping directed by Kenneth S. Kendler. The total number of participants in the VTS study was 1,428 monozygotic (MZ) female twins and dizygotic (DZ) female twins. The final sample size for our multivariate analysis after pairwise deletion for missing values was 1,100 (334 MZ twin pairs and 216 DZ twin pairs). The participants’ ages ranged from 18 to 55 years; the average age was 30 years. The twins were drawn from the population-based registry operated by the Department of Human Genetics at the Medical College of Virginia. (For a description of zygoosity diagnosis and other features of the data collection, see Kendler, Neale, Kessler, Heath, & Eaves, 1992.) Full-face color photos were taken of all twins during the interview phase of the study, to provide an independent check on similarity or dissimilarity of physical features in determining zygoosity.

**Measures**

Because of the difficulty in measuring symptoms of depression (Silberg et al., 1991; Torgersen, 1986), three measures of depression were used in the current study. The participants received a self-rating problem checklist for symptoms of depression experienced during their lifetime. Ratings were on a continuum from not at all distressed to extremely distressed. These questions were selected from those in the *Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R)*; American Psychiatric Association, 1980) for diagnosing major depression. A summary question was asked about lifetime incidence of at least three symptoms of depression that lasted at least 2 weeks (DEPA6). This checklist was filled out by the participants at least 1 year before they responded to the personal interview of the Structured Clinical Interview for *DSM-III* Diagnosis (SCID; Endicott & Spitzer, 1978; Spitzer & Williams, 1983). The interrater reliability for the *DSM-III-R* diagnosis of major depression (MD DIAGNOSIS) was assessed in 53 jointly conducted interviews. The kappa coefficient for agreement was 0.96 + 0.04, which is significant at less than the .00001 level.

Additionally, the twins responded to a self-rating problem checklist subset (SCL-D) of the SCL-54 (Derogatis, Lipman, & Covi, 1971), indicating symptoms of depression experienced in the last 30 days. The dimensions of symptoms
that the SCL-54 checklist includes are generalized anxiety, depression, somatic complaints, and phobic anxiety. VTS reliability checks ($N = 1,433$) have shown a .56 correlation between this instrument and checklists completed 1 year later.

Thus, there were three independent measures of depression taken for each twin: (a) a self-rating checklist of lifetime incidence of depression (DEPA6), (b) a structured interview with diagnostic information reviewed and diagnosed by a trained psychiatrist (MD DIAGNOSIS), and (c) a subset of a symptom checklist (SCL-D).

Procedure

As part of the interview process, we took a Polaroid (SX-70) color photograph of each participant’s upper body and face against a neutral background. The participants were asked not to smile. Photographs were taken until we had a sufficiently clear one that could be included with the completed interview data. Those twins for whom adequate photographs were not obtained mailed personal photographs to the Virginia Twin Study. Those photographs were not included in the current study if their quality was not similar to the Polaroid photographs taken by the interviewers.

Ratings of the photographs were made by four independent, Caucasian judges, balanced for age and gender (two men, one in his early 30s and the other over 50; two women, one in her mid-20s and one over 50), who were unaware of the zygosity of the twins. Ratings were made on a scale ranging from very unattractive (1) to very attractive (9). The photographs were sorted so that no two twins from the same set were rated consecutively. In contrast to the method used in a pilot study (McGovern, 1989), in this study the raters received no prior training, so that their unique perceptions of physical attractiveness might be measured. They were asked to rate grooming separately from physical attractiveness.

All four judges were checked for intrarater reliability of assessment by test–retest methods. With the large number of photos being rated, we did not expect memory for individual ratings to be significant. However, to adjust for the possibility that memory would spuriously inflate the reliability coefficient, we had the ratings done at least 1 week apart, and the photographs were in a different order. Also, three of the raters had rated a subset of 226 of the photographs for a pilot study approximately 1½ years earlier, and these sets of data allowed us to estimate test–retest reliability.

Method of Data Analysis

We used multiple regression analysis to test the relationship between physical attractiveness and symptoms of depression beyond those attributable to age. Pearson correlations were computed for continuous measures of physical attractiveness, grooming, age, and SCL-D for the whole sample and by zygosity.
Point-biserial correlations were computed for DEPA6 and MD DIAGNOSIS, which were dichotomous variables.

Statistical Hypothesis

We hypothesized that a relationship between physical attractiveness and depression would be found, with an elevation in symptom scores for very unattractive women (Farina et al., 1986; Farina et al., 1977; Napoleon et al., 1980).

Summary of Methods of Data Analysis

We used Pearson correlations to assess the reliability of the physical attractiveness measures. Each rater was compared against an average rating of attractiveness. Also, test–retest reliability was computed for three of the raters who had rated a subset of the photographs 1½ years earlier. The correlations between MZ twin pairs who shared equal genes and DZ twin pairs who shared ½ of their genes (on the average) were compared for further validation of the physical attractiveness construct.

To test the main hypothesis, we used Pearson and polyserial correlations to measure the relationship between physical attractiveness and the three measures of depression: (a) the self-rating problem checklist, (b) the interview diagnostic information, and (c) the SCL-D measure. We examined the scatter plot to check for a curvilinear relationship between physical attractiveness and symptoms of depression. Multiple regression was used to see if there was a relationship between physical attractiveness and symptoms of depression, with age differences taken into account.

Results

Physical Attractiveness

As suggested by previous research, achieving interrater reliability was not difficult. Scores ranged from 1 to 9, with mean and median scores of 5.5 (SD approximately 1.4). The correlation between each of the four judges’ ratings and the average rating of physical attractiveness was .82, .85, .86, and .82, respectively. Three of the four judges had rated a pilot sample of 226 twins 1½ years before the current rating session. The same photos were mixed in with the current set, and the raters were checked for reliability against their previous ratings for the same photos. Test–retest reliability scores were high: For Rater 1, the correlation with the prior rating was .77; for Rater 2, .83; and for Rater 3, correlated .85. Both methods indicate high reliability of ratings (see Table 1).

When the scores were broken down according to zygosity, the correlation in attractiveness ratings was .65 for identical (monozygotic) twins and .33 for fraternal (dizygotic) twins. The correlation between physical attractiveness and a
TABLE 1  
Mean Scores for Physical Attractiveness and Three Measures of Depression  
in a Female Twin Population

<table>
<thead>
<tr>
<th>Measure</th>
<th>Total sample(^a) (N = 1,100)</th>
<th>MZ twin pairs (n = 334)</th>
<th>DZ twin pairs (n = 216)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beauty(^b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M)</td>
<td>5.47</td>
<td>5.51</td>
<td>5.42</td>
</tr>
<tr>
<td>(SD)</td>
<td>1.13</td>
<td>1.34</td>
<td>1.13</td>
</tr>
<tr>
<td>MD DIAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M)</td>
<td>.30</td>
<td>.29</td>
<td>.32</td>
</tr>
<tr>
<td>(SD)</td>
<td>.46</td>
<td>.46</td>
<td>.47</td>
</tr>
<tr>
<td>DEPA6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M)</td>
<td>.55</td>
<td>.55</td>
<td>.55</td>
</tr>
<tr>
<td>(SD)</td>
<td>.50</td>
<td>.50</td>
<td>.50</td>
</tr>
<tr>
<td>SCL-D(^c)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M)</td>
<td>.57</td>
<td>.56</td>
<td>.58</td>
</tr>
<tr>
<td>(SD)</td>
<td>.10</td>
<td>.09</td>
<td>.10</td>
</tr>
</tbody>
</table>

Note. MD DIAG = psychiatrist’s diagnosis of depression. DEPA6 = self-rated lifetime incidence of depression. SCL-D = self-rated subset of the SCL-54 (Deroogatis et al., 1971).

\(^a\)Participants ranged in age from 18 to 55 years old. The mean age was 30 years old. \(^b\)Physical attractiveness scores ranged from 1 to 9. \(^c\)SCL-D scores ranged from .44 to .92. MD DIAG and DEPA6 were dichotomous variables.

measure of grooming was .72, but MZ and DZ twin correlations were not significantly different. An inverse correlation between physical attractiveness and age was found \((r = -.24)\), with older twins rated as significantly less attractive than younger twins.

Relationship Between Attractiveness and Symptom Scores

We predicted that there would be an inverse relationship between physical attractiveness and symptom scores. This hypothesis was not supported for any of the three measures of depression (SCL-D, MD DIAGNOSIS made by a physician, or a self-rating of lifetime incidence of symptoms of depression [DEPA6]; see Table 2). Scatter plots of the data did not reveal any curvilinear trends. A correlation of .00 was found between attractiveness ratings and symptoms of depression as measured by the SCL-D. The MD DIAGNOSIS of depression did not correlate with ratings of physical attractiveness \((r = .01)\), nor did self-ratings of lifetime incidence of depressive symptoms (DEPA6; \(r = .01\)). Tests for normality of the data did not reveal any significant departures. SCL-D scores ranged from 0.44 to 0.92, with a mean of 0.57 \((SD = 0.10)\).

The MD DIAGNOSIS was a dichotomous variable. It correlated with age \((r = .13)\), meaning that older twins were slightly more likely to receive a diagnosis
of depression. The DEPA6 variable was also dichotomous and did not correlate significantly with age ($r = .05$).

A check of the assumption that twins share equal environments was done for physical attractiveness by taking a measure of grooming for all twins. We know that MZ twins spend a large amount of time together, and we would expected that MZ twins would be more alike in their grooming than DZ twins. The similarity in grooming could then affect ratings of physical attractiveness. The results showed that ratings of grooming for MZ twins and DZ twins were not significantly different (see Table 3).

**Discussion**

The results of the study indicate that perceptions of physical attractiveness can be reliably measured. The hypothesis predicting an inverse relationship between physical attractiveness and symptoms of depression was not supported.

**TABLE 2**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Age</th>
<th>MD DIAG</th>
<th>DEPA6</th>
<th>SCL-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beauty</td>
<td>-.24*</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>.13*</td>
<td>.05</td>
<td>-.08</td>
</tr>
<tr>
<td>MD DIAG</td>
<td></td>
<td></td>
<td>.29*</td>
<td>.15*</td>
</tr>
</tbody>
</table>

*Note. N = 1,100. MD DIAG = psychiatrist's diagnosis of depression. DEPA6 = self-rated lifetime incidence of depression. SCL-D = self-rated subset of the SCL-54 (Derogatis et al., 1971). All coefficients are product-moment correlations based on subject deletion when the values were missing. *p < .05.

**TABLE 3**

<table>
<thead>
<tr>
<th>Twin correlation</th>
<th>MZ</th>
<th>DZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beauty$^a$</td>
<td>.65</td>
<td>.33</td>
</tr>
<tr>
<td>Grooming</td>
<td>.55</td>
<td>.47</td>
</tr>
<tr>
<td>MD DIAG$^a$</td>
<td>.47</td>
<td>.27</td>
</tr>
</tbody>
</table>

$^a$MZ and DZ twin correlations were significantly different at $p < .05$. 
The results of the study address two main areas: (a) the rating and analysis of physical attractiveness and (b) the test of the hypothesis concerning attractiveness and depression.

The interrater and test–retest reliability plus the proportionality of genetic influence strongly support the validity of the physical attractiveness construct (Patzer, 1985). The main hypothesis, which predicted an inverse relationship between level of attractiveness and symptoms of depression, was not supported. This hypothesis was based on the “beautiful is good” body of research, which supports the theory that physically attractive people are treated differently by others. According to behavioral theories of depression, this differential treatment should, in turn, affect an individual’s level of depression. In this study we did not find such a relationship; several possible explanations are suggested for this finding.

First, in all of the preceding studies wherein correlations ranging from .17 to .42 were found between physical attractiveness and depression scores, the authors noted problems with either small sample sizes or lack of methodological rigor. In addition, the higher correlations were all found in clinical settings. There may simply be no correlation between level of physical attractiveness and symptoms of depression in the general population.

Another possibility is that our results are biased. Our sample consisted of Caucasian, female twins. Although the participants were from a population-based sample, perhaps the very fact of being twins makes them different. Others have noted that being a twin can affect the relationship between social support and depression (Kessler, Kendler, Heath, Neale, & Eaves, in press). Perhaps being a twin neutralizes any effects of physical attractiveness. Also, the data collected were cross-sectional, with ages ranging from 18 to 55 years. Perhaps it is more important to be physically attractive as a child or as a young adolescent. Although interrater and test–retest reliabilities were established for the ratings of physical attractiveness, perhaps there was something unique to the four raters that may not generalize to the general population. Furthermore, others have noted the difficulty in measuring symptoms of depression (Silberg et al., 1991; Torgersen, 1986). In the current study, the three measures of depression correlated only .27, .29, and .15 with each other (see Table 2). Two of the measures evaluated neurovegetative symptoms of depression, but for different periods of time. The DEPA6 related to lifetime incidence, and the MD DIAGNOSIS related to a period of depression in the preceding year. Finally, the SCL-D assessed the cognitive dimensions of depression, such as hopelessness, worry, and lack of pleasure, within the preceding 30 days. Each measure assessed a different aspect or time period for symptoms of depression, and this way of measuring may account for the lack of high correlations; yet none of the measures correlated with physical attractiveness. Thus, care must be taken in generalizing any of the findings.

Nonetheless, with the reservations just stated kept in mind, how attractive one is may not affect level of depression. Assuming that individuals are treated
differently based on their physical attractiveness, we suggest several possible explanations for this finding.

The unique traits of each individual were not taken into account, and these traits may protect people from an unfavorable environment or a set of genes disposing one to physical unattractiveness. Future researchers could investigate the relationship between level of attractiveness and symptoms of depression by including other internal variables, such as self-concept and locus of control. For instance, objective (external and multiple) ratings of attractiveness may be less important than self-ratings of attractiveness. "Subjective" physical attractiveness represents a central dimension of self-concept across the life span (Adams, 1977; Berscheid, Dion, Hatfield (Walster), & Walster, 1971; Sorell & Nowak, 1981). Perhaps those who believe themselves to be attractive are able to offset unfavorable reactions in the environment. Or perhaps those who are depressed think of themselves as unattractive. For example, Noles, Cash, and Winstead (1985) found that depressed individuals saw themselves as less attractive than nondpressed individuals considered themselves, although the objective raters found no differences. In fact, the depressed individuals negatively distorted and the nondepressed individuals positively distorted their perceived attractiveness, compared with the objective ratings. Future researchers can investigate whether respondents rate depression as a cause or a consequence of unattractiveness.

REFERENCES


*Received May 10, 1995*