Dependent Stressful Life Events and Prior Depressive Episodes in the Prediction of Major Depression

The Problem of Causal Inference in Psychiatric Epidemiology

Kenneth S. Kendler, MD; Charles O. Gardner, PhD

Context: Most environmental risk factors for psychiatric disorders cannot be studied experimentally, making causal attributions difficult. Can we address this question by using together 2 major methods for causal inference: natural experiments and specialized statistical methods?

Objective: To determine the causal relationship between dependent stressful life events (dSLEs) and prior depressive episodes (PDEs) and major depression (MD).

Design: Assessment of risk factors and episodes of MD at interview. Statistical analyses used the co-twin control and propensity score–matching methods.

Setting: General community.

Participants: Four thousand nine hundred ten male and female twins from the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders.

Main Outcome Measure: Episodes of MD.

Results: We found that dSLEs were strongly associated with risk for MD in female (odds ratio [OR], 5.85) and male (4.55) twins in the entire sample and, at considerably lower levels, in female (2.29) and male (2.19) monozygotic twins discordant for dSLE exposure. A case-control sample matched on propensity score showed a moderate association in female (OR, 1.79) and male (1.53) twins. A PDE strongly predicted risk for MD in female (OR, 3.68) and male (5.20) twins in the entire sample. In monozygotic pairs discordant for exposure, the association was weaker in male (OR, 1.41) and absent in female (1.00) twins. A case-control sample matched on propensity score showed a moderate association between PDE and depressive episodes in male (OR, 1.58) and female twins (1.66).

Conclusions: Although dSLEs have a modest causal effect on the risk for MD, a large proportion of the observed association is noncausal. The same pattern is seen for PDEs, although the causal impact is somewhat more tenuous. For environmental exposures in psychiatry that cannot be studied experimentally, co-twin control and propensity scoring methods—which have complementary strengths and weaknesses—can provide similar results, suggesting their joint use can help with the critical question of causal inference.

Arch Gen Psychiatry. 2010;67(11):1120-1127

LUCIDATING CAUSAL PATHWAYS TO PSYCHIATRIC ILLNESS IS A CRITICAL RESEARCH GOAL. HOWEVER, WITH MANY RISK FACTORS FOR PSYCHOPATHOLOGY, CAUSAL INFERENCE IS PROBLEMATIC BECAUSE RANDOMIZED CONTROLLED TRIALS—THE GOLD STANDARD FOR CLARIFYING CAUSATION—ARE, FOR ETHICAL OR PRACTICAL REASONS, IMPOSSIBLE. TWO APPROACHES REMAIN TO EVALUATE CAUSAL PROCESSES: NATURAL EXPERIMENTS AND STATISTICAL METHODS.1-5

In this report, we examine whether dependent stressful life events (dSLEs)6-9; events likely influenced by the individual’s behavior) and prior depressive episodes (PDEs) increase risk for major depression (MD). Increased rates of MD follow dSLEs7-9 and PDEs.10-12 However, the causal nature of these associations is unclear.

The critical question is whether the association between these risk factors and MD is causal (path c in Figure 1) or arises from covariates influencing both risk factors (path a) and disease (path b). As articulated by the counterfactual/interventionist theories of causation,13-15 if the association is causal, reducing risk factor exposure will reduce rates of disease. If the association is noncausal, altering risk factor exposure will not affect illness rates.

Many factors predispose to both dSLEs and MD, including neuroticism, PDEs, and genetic risk for MD.16-21 Thus, the association between dSLEs and MD might result from these shared covariates. Because numerous risk factors for MD—including...
genes and childhood adversities—have long-lasting effects. The tendency for MD to recur could arise without one depressive episode having a causal effect on the next.

In this report, we address these questions using a natural experiment (twins) and a specialized statistical method (propensity analysis). The co-twin control method examines outcomes in twin pairs discordant for risk factor exposure. Twin pairs raised together are matched for their rearing environment and their genes (partly for dizygotic [DZ] and completely for monozygotic [MZ] pairs) although not for environmental experiences unique to each twin.

Propensity analysis simulates a case-control study in a cohort assessed for the risk factor, outcome, and covariates predicting risk factor exposure. For each case (an individual exposed to the risk factor), an unexposed control is selected matched on the probability of being a case as predicted by the covariates (ie, the propensity to risk factor exposure). The co-twin control and propensity analyses attempt to isolate the impact of single risk factors and discriminate their effect from the background context in which these risk factors arise. We applied these methods separately in male-male and female-female twin samples for the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders. Participants derive from 2 related studies in white same-sex twin pairs from the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders. Participants were ascertained from the birth-certificate–based Virginia Twin Registry. Female-female twin pairs, born 1934 through 1974, were eligible if both members responded to a mailed questionnaire in 1987 through 1988. We use data from all 4 waves (FF1, FF2, FF3, and FF4), which included 85% to 92% of eligible twins. Mean interwave intervals were 17.3 months for FF1 to FF2, 45.0 months for FF2 to FF3, and 36.1 months for FF3 to FF4. Male-male pairs, born 1940 through 1974, were ascertained from registry records by a telephone interview (MM1), with a 72% response rate, and followed up with a face-to-face interview (MM2), with an 83% response rate. The mean interwave interval was 19.0 months. Zygosity was determined by a face-to-face interview (MM2), with an 83% response rate. The mean interwave interval was 19.0 months. Zygosity was determined by discriminate function analyses using “twin questions” validated against DNA genotyping. The mean (SD) age and education of the twins were 36.3 (8.2) and 14.3 (2.2) years, respectively, at the MM2 interview and 37.0 (9.1) and 13.6 (2.6) years, respectively, at the MM2 interview.

This project was approved by the human subject committees at Virginia Commonwealth University. Written informed consent was obtained before the MM2 interviews, and verbal consent was obtained before the MM1 interviews.

MEASUREMENTS

We assessed the occurrence, to the nearest month during the year before the MM2 interview, of 11 classes of personal and 4 classes of network events. Interrater reliability (κ) values for the occurrence and dating of our SLE categories were 0.93 and 0.82, respectively.

Dependence of SLEs, conceptualized as the probability that a respondent’s own behavior contributed to the SLE, was interrater rated on a 4-point scale as clearly independent, probably independent, probably dependent, and clearly dependent. Interrater reliability was assessed from tape recordings of 92 randomly selected SLEs. Test-retest reliability was obtained by blindly reinterviewing 191 respondents at a mean interval of 4 weeks. Using the Spearman correlation (r) and weighted kappa (κ), we determined the test-retest and interrater reliabilities to be r = 0.77/κ = 0.63 and r = 0.89/κ = 0.79, respectively. We defined a dSLE as probably or clearly dependent. Diagnoses of MD in the past year were made according to DSM-IV criteria. For each reported episode, the respondents described its duration and date of onset and offset. Of the MD episodes examined in our dSLE and PDE analyses, 36.4% and 40.4%, respectively, denied prior episodes (ie, were potentially first onsets).

We explored the relationship between dSLEs and risk for depressive episode onset in the month of the dSLE and 2 subsequent months in the FF4 and MM2 interviews. For our propensity analysis, we chose 18 covariates as predictors of dSLE exposure. Beginning with an extensive list of variables used in 2 prior comprehensive analyses of the etiology of MD in this sample, we selected variables that were assessed before the FF4 and MM2 interviews or were collected to reflect events occurring before that interview. These variables (and, where available, the original scales and publications showing their association with MD) were (1) birth year, (2) mean parental warmth measured by the Parental Bonding Instrument, (3) childhood sexual abuse, (4) parental loss due to death or divorce in childhood, (5) neuroticism measured by the short version of the Eysenck Personality Questionnaire, (6) extraversion (extraversion reversed) measured by the short version of the Eysenck Personality Questionnaire, (7) self-esteem measured by the Rosenberg scale, (8) early-onset anxiety disorder, (9) conduct disorder, (10) lifetime trauma, (11) social support, (12) lifetime diagnosis of alcohol abuse or dependence, (13) lifetime diagnosis of nicotine dependence, (14) lifetime diagnosis of illicit drug abuse or dependence, (15) history of divorce, (16) history of MD more than 1 year before the FF3 or MM1 interview, (17) marital quality (high risk is defined as being married in the bottom 20% of quality); medium risk, unmarried; and low risk, married in the top 80% of quality), (18) difficulties, described as the number of life events in the past year outside the 3-month window before the depressive episode onset for cases (or chosen at random for controls).

To evaluate the relationship between PDE and risk for a subsequent depressive episode, MD episodes reported at the MM1 and FF1 interviews were used as risk factors, and those reported at the MM2 and FF2 interviews were used as the outcomes. We began with the same possible list of covariates, but herein, to avoid possible bias, our final list was shorter because it had to reflect events or risk factors present before the

**Figure 1.** The interrelationship among a risk factor, a potential covariate, and an outcome. The causal interrelationship between the risk factor and the outcome is depicted by path c, whereas the noncausal pathway is captured by paths a and b.
We wished to compare propensity matching to the co-twin control approach. To understand the potential success of our approach, we examined covariate imbalance, that is, the degree to which our groups (general population, discordant twins, and our propensity score–constructed case-control sample) were comparable to which our groups (general population, discordant twins, and our propensity score–constructed case-control sample) were comparable to our groups (general population, discordant twins, and our propensity score–constructed case-control sample). We calculated odds ratios (ORs) using standard logistic regression as operationalized in the SAS procedure. We repeated all analyses in our matched samples using conditional logistic regression and found very similar results.

For our co-twin control analyses, we obtained twin pairs discordant for the risk factor exposure, separately by zygosity group. Only same-sex pairs were used. Male-male and female-female samples were analyzed separately.

We calculated odds ratios (ORs) using standard logistic regression as operationalized in the SAS procedure. We repeated all analyses in our matched samples using conditional logistic regression and found very similar results. We wished to compare propensity matching to the co-twin control approach. To understand the potential success of our approach, we examined covariate imbalance, that is, the degree to which our groups (general population, discordant twins, and our propensity score–constructed case-control sample) were matched on our determined covariates. Covariate imbalance was assessed by standardizing the sample around the grand mean for each covariate and then calculating means for each group. Consistent with previous approaches, we defined covariate imbalance as a difference of 0.25 SD or more in the mean of the standardized covariate in the 2 groups.

In our propensity score analyses, we used a nearest neighbor or a greedy matching algorithm implemented in an SAS macro to create matched pairs discordant for exposure. This algorithm examines cases sequentially and finds the closest match among the controls on the basis of propensity score (the probability of being a case from multivariate logistic regression). On the first iteration, a match is retained if propensity scores are within 0.00001. The matched pair is then removed from the data set and not considered further. Additional iterations allow matches within 0.0001, 0.001, 0.01, and 0.1. If a match has not been found at that point, the case is excluded. So that we could include individuals with missing values for some variables, we used a multiple imputation approach. The method used was that of Raghunathan et al.,49 as implemented in IVEware software. We created 10 imputed data sets for analysis, and the results were combined with estimates and standard errors calculated as described by Rubin31 and Li et al.32

The mean results from these 4 regression analyses predicting dSLEs and PDEs in male and female twins are given in eTable 1 and eTable 2 (http://www.archgenpsychiatry.com).

We report the C statistic from the co-twin control and propensity analyses; this reflects the degree to which exposed and unexposed individuals in the initial population differ on their propensity score. A C value of 0.5 indicates no predictive power (ie, a correct prediction half the time). A C value of 1.0 means perfect prediction, in which cases and controls reflect entirely separate distributions. For optimal propensity score matching, a moderate overlap of the 2 distributions (eg, C values of approximately 0.7) is ideal. At high C values, it becomes increasingly difficult to match cases and controls because many cases have covariate values outside the range observed in controls.

Confidence intervals and P values are based on profile likelihoods. Because of clear a priori expectations that dSLEs and PDEs should increase the risk for MD, we present 1-tailed P values and 90% confidence intervals (CIs).

### RESULTS

#### dSLEs AS A CAUSAL RISK FACTOR FOR AN MD EPISODE

**Male-Male Pairs**

We began with a sample of 2394 members of male-male twin pairs, including both members of 706 MZ and 491 DZ pairs. Of these, 496 were exposed to a dSLE in the year before the MM1 interview. Exposure to a dSLE was strongly associated with an onset of MD within 3 months (OR, 4.55; 90% CI, 3.49-5.90; P < .001) (Figure 2). Participants in this sample who were vs were not exposed to the dSLE were poorly matched, differing substantially on a number of the covariates selected to reflect risk factors for dSLEs (Table 1). The 2 groups displayed covariate imbalance on the following 10 variables: parental warmth, neuroticism, conduct disorder, lifetime traumas, alcohol abuse/dependence, illicit drug abuse/dependence, history of MD before the last year, marital quality, difficulties, and birth year.

This sample contained 312 twin pairs discordant for dSLE exposure, of whom 133 were DZ and 179 were MZ. Compared with observations in the entire sample, the association between dSLE exposure and depressive episode onset was still statistically significant but considerably lower in magnitude in discordant DZ pairs (OR, 3.31; 90% CI, 1.54-7.84; P = .004) and lower still in discordant MZ pairs (2.19; 1.23-3.93; P = .01). As seen in Table 1, the DZ pairs discordant for dSLE exposure were better matched than the entire sample but still had covariate imbalance on the following 6 variables: parental...
warmth, neuroticism, self-esteem, history of divorce, history of MD before the past year, and marital quality. The MZ pairs discordant for dSLE exposure were much better matched than the 2 previous groups, having a covariate imbalance on only the variable of marital quality.

In our propensity analysis, we obtained a mean C statistic across the 10 imputations of 0.711, which indicates a good degree of overlap in propensity scores between the exposed and unexposed groups. As expected, we did well with our matching, obtaining a mean of 491.9 controls for our 496 cases. The case-control sample created by our propensity analysis had no covariate imbalance. The OR between dSLE exposure and a depressive outcome in this analysis was 1.53 (90% CI, 1.24-1.88; P < .001).

Female-Female Pairs

We began with a sample of 1938 members of female-female twin pairs, which included both members of 503 MZ and 326 DZ twins as well as 3 twin pairs of unknown zygosity and 274 twins, without their co-twin. Of this sample, 155 had been exposed to a dSLE in the year before the FF4 interview. In this entire sample, exposure to a dSLE was strongly associated with an onset of MD within 3 months (OR, 5.85; 90% CI, 4.22-8.06; P < .001). Exposed and unexposed participants were again poorly matched, demonstrating covariate imbalance on the following 8 variables: parental loss, conduct disorder, lifetime traumas, alcohol abuse/dependence, illicit drug abuse/dependence, marital quality, difficulties, and birth year.

This sample contained 108 twin pairs discordant for dSLE exposure, of whom 46 were DZ and 62 were MZ. Compared with that seen in the entire sample, the association between dSLE exposure and depressive episode onset was considerably lower but remained significant in discordant DZ pairs (OR, 3.23; 90% CI, 1.30-8.85; P = .02) and lower still in discordant MZ pairs (2.29; 1.02-5.44; P = .05). The DZ pairs discordant for dSLE exposure had covariate imbalance on the following 3 variables: self-esteem, illicit drug abuse/dependence, and marital quality. The MZ pairs discordant for dSLE exposure were even better matched, with covariate imbalance only on marital quality.

Our propensity analysis achieved a mean C statistic of 0.713, indicating a substantial degree of overlap in propensity scores between cases and controls. As a result, we were able to obtain matched controls for a mean of 153.9 of 155 cases. The selected case-control sample demonstrated no covariate imbalance. The OR between dSLE exposure and a depressive outcome in this sample was 1.79 (90% CI, 1.33-2.41; P = .001).

PDEs AS CAUSAL RISK FACTORS FOR SUBSEQUENT DEPRESSIVE EPISODES

Male-Male Pairs

We began with a sample of 2908 members of male-male twin pairs personally assessed at our MM2 interview, 162 of whom reported 1 or more PDEs in the year before our MM1 interview. In this sample, exposure to a PDE in the year before the MM1 interview was strongly associated with the occurrence of 1 or more depressive episodes in the year before the MM2 interview (OR, 5.20; 90% CI, 3.67-7.27; P < .001). As seen in Table 2, exposed and unexposed twins in this sample were poorly matched, having covariate imbalance on the following 10 variables: a history of MD, neuroticism, childhood sexual abuse, social support, self-esteem, marital quality, SLEs, difficulties, parental warmth, and birth year.

This sample contained 119 twin pairs discordant for exposure to a PDE, of whom 51 were DZ and 68 were MZ. Compared with that observed in the entire sample, the association between PDEs and depressive episode onset was considerably lower in discordant DZ pairs (OR, 2.31; 90% CI, 0.96-5.91; P = .06) and lower still in discordant MZ pairs (1.41; 0.64-3.14; P = .24). The DZ and MZ pairs discordant for a PDE had covariate imbalance on 5 variables (Table 2). For DZ pairs, these variables were neuroticism, self-esteem, marital quality, SLEs, and difficulties. For MZ pairs, these were a history of MD, neuroticism, self-esteem, marital quality, and SLEs.

Our propensity score analysis produced a mean C statistic of 0.840, which is higher than ideal. However, we were successful in obtaining matched controls for a mean of 160.3 of 162 cases. No covariate imbalance was found in these cases and controls (Table 2). The OR between
was 1.58 (90% CI, 0.87-2.88; OR, 3.68; 90% CI, 2.66-5.04; P = .001). Exposed and unexposed twins were poorly matched, with covariate imbalance on 4 variables: a history of MD, neuroticism, self-esteem, and marital quality. When we examined DZ and MZ twin pairs discordant for dSLE exposure, the association with MD fell substantially but remained statistically significant in both groups. As expected, the OR was lower in the MZ than in the DZ pairs given that the former controls entirely for genetic risk factors. In participants matched for family background and partially (DZ pairs) or completely (MZ pairs) for genes, dSLE remained predictive of depressive episodes. Of interest, MZ twin pairs discordant for dSLE exposure were very similar on our covariates, differing in male and female twins only in marital quality.

Our propensity method worked as expected in obtaining controls well matched to our cases. The OR observed in the propensity score analysis was similar and statistically significant in both male and female twins but also substantially lower than that observed in the general population.

Although the OR produced by the propensity score method was lower in both male and female twins than that obtained using the co-twin control method in MZ pairs, these 2 estimates of causal effect had widely overlapping CIs in both sexes. For male and female twins, the CIs for our 2 best estimates of causal effect overlapped in the range of 1.2 to 1.9 and 1.3 to 2.4, respectively. In both sexes, these 2 methods have produced average across our 10 iterations, matched controls for 170.2 of the 184 cases, with no covariate imbalance between the 2 groups. The OR between exposure to a PDE and a subsequent depressive episode in this analysis was 1.66 (90% CI, 0.95-2.89; P = .07).

| Table 2. Standardized Differences in the Means of Covariates in Male Individuals Exposed and Unexposed to a Prior Depressive Episode* |
|------------------|------------------|------------------|------------------|
| Covariate        | Total Sample     | Discordant DZ Twins | Discordant MZ Twins | Propensity Score–Matched Sample |
| Years of education | 0.172            | 0.028             | 0.072             | 0.053                          |
| Mean parental warmth | 0.358            | 0.232             | 0.090             | 0.058                          |
| Childhood sexual abuse | 0.328            | 0.166             | 0.249             | 0.066                          |
| Neuroticism | 1.078            | 0.704             | 0.728             | 0.051                          |
| Introversion | 0.083            | 0.110             | 0.038             | 0.068                          |
| Self-esteem | 0.669            | 0.698             | 0.498             | 0.057                          |
| Social support | 0.289            | 0.180             | 0.166             | 0.076                          |
| Marital quality | 0.483            | 0.408             | 0.464             | 0.039                          |
| SLEs | 0.849            | 0.490             | 0.438             | 0.055                          |
| History of MD before the past year | 0.553            | 0.244             | 0.324             | 0.031                          |
| Difficulties | 0.629            | 0.414             | 0.064             | 0.069                          |
| Birth year | 0.311            | 0.000             | 0.000             | 0.069                          |

Abbreviations: DZ, dizygotic; MD, major depression; MZ, monozygotic; SLEs, stressful life events.

*Scores with a value of at least 0.25, which meet the definition of a standardized covariate in the 2 groups, are given in boldface type. The propensity score reflects the mean of 10 iterations.

PDE exposure and a depressive outcome in this analysis was 1.58 (90% CI, 0.87-2.88; P = .10).

Female-Female Pairs

We began with a sample of 2002 members of female-female twin pairs personally assessed at our FF2 interview, 184 of whom reported 1 or more PDEs in the year before our FF1 interview. In that sample, exposure to a PDE was strongly associated with the occurrence of 1 or more PDEs in the year before the FF1 interview (OR, 3.68; 90% CI, 2.66-5.04; P = .001). Exposed and unexposed twins were poorly matched, with covariate imbalance on 4 variables: a history of MD, neuroticism, social support, self-esteem, marital quality, SLEs, difficulties, and parental warmth.

This sample contained 132 twin pairs discordant for exposure to a PDE, of whom 63 were DZ and 69 were MZ. Compared with the results in the entire sample, the association between a PDE and a subsequent depressive episode was considerably lower in discordant DZ pairs (OR, 2.04; 90% CI, 0.97-4.47; P = .06) and disappeared entirely in discordant MZ pairs (1.00; 0.45-2.23; P = .99). The DZ and MZ pairs discordant for a PDE in the year before our FF1 interview had covariate imbalance on 4 variables: a history of MD, neuroticism, self-esteem, and SLEs for DZ pairs and neuroticism, self-esteem, marital quality, and SLEs for MZ pairs.

Our propensity analysis produced a mean C statistic of 0.857, again indicating potential problems due to the limited overlap of the propensity score distribution of the exposed and unexposed twins. Indeed, we obtained, on average across our 10 iterations, matched controls for 170.2 of the 184 cases, with no covariate imbalance between the 2 groups. The OR between exposure to a PDE and a subsequent depressive episode in this analysis was 1.66 (90% CI, 0.95-2.89; P = .07).

**COMMENT**

We specifically studied dSLEs because their causal relationship with psychiatric illness is inherently unclear and their association with MD is typically stronger than that observed for independent SLEs. Indeed, the occurrence of dSLEs was strongly related to the risk for MD episodes in our entire sample in both male and female twins. In our detailed set of covariates, those exposed to dSLEs (cases) frequently differed from those who were unexposed (controls), immediately raising concern that the association between dSLEs and subsequent depressive episodes may not be entirely causal and may instead be at least partly mediated via these covariates along paths a and b in Figure 1.

When we examined DZ and MZ twin pairs discordant for dSLE exposure, the association with MD fell substantially but remained statistically significant in both groups. As expected, the OR was lower in the MZ than in the DZ pairs given that the former controls entirely for genetic risk factors. In participants matched for family background and partially (DZ pairs) or completely (MZ pairs) for genes, dSLE remained predictive of depressive episodes. Of interest, MZ twin pairs discordant for dSLE exposure were very similar on our covariates, differing in male and female twins only in marital quality.

Our propensity method worked as expected in obtaining controls well matched to our cases. The OR observed in the propensity score analysis was similar and statistically significant in both male and female twins but also substantially lower than that observed in the general population.

Although the OR produced by the propensity score method was lower in both male and female twins than that obtained using the co-twin control method in MZ pairs, these 2 estimates of causal effect had widely overlapping CIs in both sexes. For male and female twins, the CIs for our 2 best estimates of causal effect overlapped in the range of 1.2 to 1.9 and 1.3 to 2.4, respectively. In both sexes, these 2 methods have produced average across our 10 iterations, matched controls for 170.2 of the 184 cases, with no covariate imbalance between the 2 groups. The OR between exposure to a PDE and a subsequent depressive episode in this analysis was 1.66 (90% CI, 0.95-2.89; P = .07).
broadly similar results showing a moderate causal impact of dSLEs on risk for MD.

**PDEs as causal risk factors for subsequent depressive episodes**

The PDEs recorded at our FF1 interviews strongly predicted risk for depressive episodes assessed at our FF2 interviews. In our set of covariates, those exposed to PDEs differed even more greatly than was seen with the dSLEs. Given these results and the fact that the risk factor and outcome variables were the same (ie, both are episodes of MD) and should share most predictors, the a priori probability that the observed association might be substantially noncausal is particularly high.

Indeed, when we examined DZ and MZ twin pairs discordant for PDE exposure, the association with MD fell sharply from that seen in the general population. None of the 4 resulting analyses (2 twin types in male and female twins) achieved statistical significance, although 2 of them produced trends suggesting a causal relationship. In participants matched completely for family background and partially (DZ pairs) or completely (MZ pairs) for genes, the evidence that PDEs are causally related to future MD episodes is modest and statistically inconclusive. However, MZ twins discordant for exposure to PDEs differed from each other on our covariates much more than those pairs discordant for dSLE exposure.

Our propensity method worked well in obtaining matched controls for our exposed cases and produced ORs that were modest and similar in male and female twins (1.6-1.7). Both results approached statistical significance, although 2 of them produced trends suggesting a causal relationship. In participants matched completely for family background and partially (DZ pairs) or completely (MZ pairs) for genes, the evidence that PDEs are causally related to future MD episodes is modest and statistically inconclusive. However, MZ twins discordant for exposure to PDEs differed from each other on our covariates much more than those pairs discordant for dSLE exposure.

The PDEs recorded at our FF1 interviews strongly predicted risk for depressive episodes assessed at our FF2 interviews. In our set of covariates, those exposed to PDEs differed even more greatly than was seen with the dSLEs. Given these results and the fact that the risk factor and outcome variables were the same (ie, both are episodes of MD) and should share most predictors, the a priori probability that the observed association might be substantially noncausal is particularly high.

Indeed, when we examined DZ and MZ twin pairs discordant for PDE exposure, the association with MD fell sharply from that seen in the general population. None of the 4 resulting analyses (2 twin types in male and female twins) achieved statistical significance, although 2 of them produced trends suggesting a causal relationship. In participants matched completely for family background and partially (DZ pairs) or completely (MZ pairs) for genes, the evidence that PDEs are causally related to future MD episodes is modest and statistically inconclusive. However, MZ twins discordant for exposure to PDEs differed from each other on our covariates much more than those pairs discordant for dSLE exposure.

Our propensity method worked well in obtaining matched controls for our exposed cases and produced ORs that were modest and similar in male and female twins (1.6-1.7). Both results approached statistical significance, although 2 of them produced trends suggesting a causal relationship. In participants matched completely for family background and partially (DZ pairs) or completely (MZ pairs) for genes, the evidence that PDEs are causally related to future MD episodes is modest and statistically inconclusive. However, MZ twins discordant for exposure to PDEs differed from each other on our covariates much more than those pairs discordant for dSLE exposure.

Overall, the results of the propensity analysis of the PDE-MD relationship were somewhat stronger than those obtained from the co-twin control analyses. From both analyses, it is clear that a large proportion of the substantial association between PDEs and subsequent depressive episodes is not causal. However, a direct causal effect probably exists. This interpretation is consistent with previous analyses in this and other samples in which, in the presence of many covariates, PDEs modestly predict risk for future depressive episodes.

**Strengths and limitations of the co-twin control and propensity matching methods for causal inference**

**Strengths**

Does our discordant twin or our propensity analysis provide a more accurate picture of causal processes? Critically, their strengths and limitations differ. The co-twin control method provides excellent matching for all shared environmental exposures and, for MZ twins, all genes. All risk factors falling into these 2 categories, including those we do not know about, are well controlled. However, environmental experiences unique to 1 member of the pair are not controlled for by this method. If such experiences influence the risk factors and the outcome, the co-twin control method could overestimate causal effects.

If we knew of and measured all the relevant factors that influence risk exposure, the propensity method would provide accurate and unbiased estimates of causal effects. Although this is not realistic for any behavioral outcome, our longitudinal twin cohort has a particularly rich set of depression-related risk factors. If covariates that influence risk factor exposure and outcome are left out of propensity matching, this could also result in an overestimation of causal effects. However, propensity matching can also overcontrol for variables and thereby downwardly bias causal effects. Covariates have to be causes and not consequences of risk factor exposure. If assessment of a covariate is influenced by risk factor exposure, this could produce an underestimation of the causal effects of the risk factor. For example, our propensity analysis of dSLEs used neuroticism that was measured before dSLE exposure. If we had measured it after dSLE exposure, it could reflect the reaction to the event. Inclusion as a covariate would then downwardly bias our estimates of the causal effect of dSLEs on MD.

We have somewhat more confidence in the results of our analyses of dSLE than of PDEs. Dependent SLEs are, at least in part, exogenous risk factors that proved relatively easy to control for, as demonstrated by the very low rate of covariate imbalance in our discordant MZ twins. By contrast, PDEs are a more endogenous variable, reflecting internal vulnerabilities to depressive illness. As might be expected, this proved harder to control for, as evidenced by the larger number of covariate imbalances seen in our MZ twins discordant for PDEs.

Our confidence in scientific results should increase when similar answers are obtained by different methods, especially when these 2 methods have different strengths and weaknesses, for example, in the congruent evidence of genetic influences on schizophrenia from twin and adoption studies. The present report is in this situation. The co-twin control and propensity analyses each suggest that dSLEs have a moderate true causal effect on the risk for MD. Both methods also agree broadly with the more tenuous evidence of a modest causal effect on risk for MD of PDEs. The congruence of these findings substantially increases the probability of their veracity.

**Limitations**

These results should be interpreted in the context of 2 potentially important methodological limitations. First, despite our large sample size, power was limited, especially for the discordant twin analyses and with the rarer risk factor of PDEs. Low power may contribute to the lack of more definitive findings.

Second, the co-twin control and propensity analyses assess the impact of a risk factor given the background exposures of the study population. If a risk factor is causally potent only in the presence of an uncommon background factor, it will typically produce a weak overall effect.
CONCLUSIONS

Clarifying causal processes is problematic in psychiatric epidemiology, in which controlled trials are often impossible. We examined 2 typical problems (the relationships between dSLEs and PDEs and MD) in which strong associations are present but causal processes are much harder to elucidate. Applying examples of the 2 major available approaches (twins as a natural experiment and propensity scoring as a statistical method), we obtained reassuringly similar answers. Although the observed associations were largely noncausal, a moderate causal effect of dSLEs on MD was convincingly demonstrated, whereas the causal impact of PDEs was somewhat weaker and more tentative. Natural experiments and statistical methods, especially when used together in carefully collected samples, can provide substantial help in clarifying the nature of the causal pathways to psychiatric illness.

Submitted for Publication: January 12, 2010; final revision received March 5, 2010; accepted June 8, 2010.

Correspondence: Kenneth S. Kendler, MD, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University School of Medicine, PO Box 980126, Richmond, VA 23298-0126 (kendler@vcu.edu).

Author Contributions: Dr Kendler had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Financial Disclosure: None reported.

Funding/Support: This study was supported in part by grants MH-068643, DA-011287, and MH-49492 from the National Institutes of Health (NIH). The Mid-Atlantic Twin Registry (MATR), of which the Virginia Twin Registry is now part, has received support from the NIH, the Carman Trust, and the W. M. Keck, John Templeton, and Commonwealth University School of Medicine, PO Box 980126, Richmond, VA 23298-0126 (kendler@vcu.edu).


Additional Contributions: Linda Corey, PhD, provided assistance with the ascertainment of twins from the Virginia Twin Registry. Judy Silberg, PhD, directs the MATR. Carol Prescott, PhD, contributed to the design and implementation of this study. Kate Lapane, PhD, provided helpful advice on the propensity score analyses.

REFERENCES


