Published in final edited form as:

Soc Psychiatry Psychiatr Epidemiol. 2004 February ; 39(2): 126–132. doi:10.1007/s00127-004-0714-z.

Are higher rates of depression in women accounted for by differential symptom reporting?

Hillary R. Bogner, MD, MSCE and Joseph J. Gallo, MD, MPH

Dept. of Family Practice & Community Medicine, University of Pennsylvania, 3400 Spruce Street/2 Gates Building, Philadelphia, PA 19104 USA, Tel.: +1-215/615-0851, Fax: +1-215/662-3591

Hillary R. Bogner: bogner@mail.med.upenn.edu; Joseph J. Gallo:

Abstract

Background—The gender difference in prevalence and incidence rates of depression is one of the most consistent findings in psychiatric epidemiology. We sought to examine whether any gender differences in symptom profile might account for this difference in rates.

Method—This study was a population-based 13-year follow-up survey of community-dwelling adults living in East Baltimore in 1981. Subjects were the continuing participants of the Baltimore Epidemiologic Catchment Area Program. Participants interviewed between 1993 and 1996 with complete data on depressive symptoms and covariates were included ($n = 1,727$). We applied structural equations with a measurement model for dichotomous data (the MIMIC – multiple indicators, multiple causes – model) to compare symptoms between women and men, in relation to the nine symptom groups comprising the diagnostic criteria for major depression, adjusting for several potentially influential characteristics (namely, age, self-reported ethnicity, educational attainment, marital status, and employment).

Results—There were no significant gender differences in the self-report of depression symptoms even taking into account the higher level of depressive symptoms of women and the influence of other covariates. For example, women were no more likely to endorse sadness than were men, as evidenced by a direct effect coefficient that was not significantly different from the null [adjusted estimated direct effect of gender on report of sadness = 0.105, 95% confidence interval (−0.113, 0.323)].

Conclusions—Men and women in this community sample reported similar patterns of depressive symptoms. No evidence that the presentation of depressive symptoms differs by gender was found.

Keywords

depression; gender differences; symptoms

Introduction

Many decades of epidemiologic studies on depression have found that women are more likely to meet criteria for major depression than men, whether one examines prevalence or incidence estimates (Gallo et al. 1993; Nolen-Hoeksema 1990; Weissman and Klerman 1992). Starting in early adolescence, more girls than boys begin to become depressed, and this gender difference in depression persists throughout adulthood (Nolen-Hoeksema 1990). The prevalence of depression among women is approximately twice the prevalence of depression

among men (Kessler et al. 1993; Robins and Regier 1991). The female preponderance in depression is present across many different countries and cultures (Weissman et al. 1996). Many different explanations have been proposed to explain this striking gender difference in the rate of depression, but the reasons for the greater vulnerability of women to depression remain incompletely understood (Nolen-Hoeksema 1987; Nolen-Hoeksema 1990).

A number of studies have examined both biological and psychosocial theories for why women are more prone to depression than are men (Nolen-Hoeksema 1987; Nolen-Hoeksema 1990; Weissman and Klerman 1977). Much less attention has been paid to gender differences in depressive symptomatology. Some of this research has raised the concern that men might be less likely to report some or all of the symptoms of depression (Nolen-Hoeksema 1987). Therefore, the gender difference in depression might be an artifact of the differences in the willingness of men and women to report specific symptoms of depression. In other words, depressed men might not reach the diagnostic threshold because they do not report the required number of associated symptoms. Although men and women tend to report balanced rates of depressed mood, some researchers have found that women appear to be more likely to present with increased appetite and weight gain as well as anxiety and somatic symptoms. For example, Frank and coworkers found that women outpatients had more appetite and weight increase, hypochondriasis, somatic anxiety, and expressed anger and hostility while men outpatients had more weight loss (Frank et al. 1988). Young and colleagues also found more appetite increase and weight gain in women outpatients (Young et al. 1990). Angst and Dobler-Mikola found that depressed women from a community sample experienced not only more appetite or weight changes, but also more sleep disturbance and feelings of worthlessness or guilt than depressed men (Angst and Dobler-Mikola 1984). Williams and coworkers found depressed women in a primary care sample reported more appetite changes and feelings of failure and remorse (Williams et al. 1995). Most recently, Kornstein and colleagues found female outpatients with chronic major depression or double depression have more sleep changes, psychomotor retardation, and somatization than men (Kornstein et al. 2000).

Our investigation differs in several ways from prior work on gender differences in depression. Our sample derives from a large ethnically diverse sample of community-dwelling adults and was not restricted to adolescents (Compas et al. 1997), specialty or primary care settings (Williams et al. 1995; Young et al. 1990), or persons who meet standard criteria for major depression (Frank et al. 1988; Kornstein et al. 2000; Young et al. 1990). Comparison of symptoms in samples derived from the health services may be distorted by factors related to help-seeking or referral patterns.

New methods have been developed in educational and psychological testing that may be of value in comparing symptoms of depression according to gender (Gallo et al. 1994; Hambleton et al. 1991; Horn and McArdle 1992; Petty and Nasrallah 1981; Schaeffer 1988; van der Linden and Hambleton 1997). In educational testing, differential item functioning, or bias, occurs when an item is less likely to be answered correctly by one group of examinees, say girls compared to boys, even when examinees at the same level of ability are compared (Muthén 1988b). In psychology, comparison of symptom endorsement across groups should account for the level of depression and potentially influential covariates (Duncan-Jones et al. 1986). Our analysis takes advantage of a measurement model with structural equations, discussed in detail below, to examine whether women differentially express symptoms such as sadness that make them more likely than men to meet criteria for depression.

To address this question, we employed symptom-level data from adults in a large community sample, the Baltimore Epidemiologic Catchment Area (ECA) sample. Prior studies from the Baltimore ECA have reported one-year prevalence estimates for affective disorders among women of 5% and among men of 2.3% (Robins and Regier 1991), but no previous study from

the ECA has investigated gender differences in symptoms reported that might account for differences in rates. In the Subjects and methods section, we discuss a model that has been applied to similar problems to compare symptoms of depression according to age (Gallo et al. 1994; Gallo et al. 1999), ethnicity (Gallo et al. 1998), and health services setting (Suh and Gallo 1997), to examine responses to the Mini-Mental State Examination for bias according to level of educational attainment (Jones 1997), and to examine differential item functioning in responses to activities of daily living/instrumental activities of daily living tasks according to age and gender (Fleishman et al. 2002).

In this investigation, we focus the model on gender differences in reporting the symptoms of depression. Specifically, we hypothesized that if we found that sadness, a key symptom of major depression, was differentially reported more by women than by men, this may account in part for the differential in rates of depression according to gender. Our goal was to uncover any male-female symptom differential reporting that might emerge in order to achieve a more comprehensive understanding of the basis for gender differences in rates of depression. In other words, we sought evidence that women might be more likely to differentially report some symptoms, such as sadness or appetite disturbance, that might lead to higher prevalence estimates among women.

Subjects and methods

The Epidemiologic Catchment Area Program

The Epidemiologic Catchment Area (ECA) Program was a coordinated set of epidemiologic surveys carried out by collaborators between 1980 and 1984 at five university-based sites in the United States. After probability sampling of adult household residents, trained lay interviewers applied a version of the Diagnostic Interview Schedule (DIS) (Robins et al. 1981). Details of the study design of the ECA have been published elsewhere (Eaton and Kessler 1985; Robins and Regier 1991).

The Baltimore ECA follow-up

In Baltimore and other ECA sites, the initial cohort aged 18 years and older at their first interview was the target for re-interview one year later; at the Baltimore site, the entire cohort of 3,481 persons surveyed at the first interview also was the target for tracing 13 years later. The participants gave permission for future follow-up at the baseline interview, and the protocol was reviewed and approved by the Committee on Human Research of The Johns Hopkins University School of Hygiene and Public Health. Prior analyses have shown that loss to follow-up was not strongly associated with depression at initial interview (Badawi et al. 1999; Eaton et al. 1992).

Analytic strategy

Depression symptom groups—At follow-up, trained lay interviewers applied the Composite International Diagnostic Interview (CIDI) (Wittchen 1994), similar to the DIS that was used in the initial interview. The standardized questions of the CIDI depression module were keyed to nine individual criteria of the case definition for major depression in the DSM-IV (American Psychiatric Association 1994). Positive responses to the questions were followed by further inquiry as to whether the symptom might be ascribed to the effects of medication, drugs, alcohol, or a physical illness. For each symptom, information on onset and recency was also gathered, so that data on recency of individual symptoms are available for the criteria of DSM-IV major depression (Von Korff and Anthony 1982).

Item parcels were created by counting the symptom group as present if any one of the symptoms forming the criterion were present within 6 months of interview (West et al. 1995). We selected

6 months as a reasonable time frame for the recall of symptoms. It was a compromise between a length of time for which recall might be unreliable (e. g., 1 year or more) and an interval that is so short (e. g., 1 month) as to result in a sparse response set. The item for “sadness,” however, represents a single symptom question, not an item parcel. The question regarding depressed affect read: “In your lifetime, have you ever had 2 weeks or more when nearly every day you felt sad, blue, depressed?”. Other symptom criterion groups were assessed in relation to more than one symptom (appetite disturbance, 4 items; sleep disturbance, 3 items; fatigue, 2 items; psychomotor change, 3 items; loss of interest, 3 items; worthlessness, 4 items; trouble concentrating, 3 items; and thoughts of death or suicide, 4 items).

Other variables under study—The MIMIC model, described below, included dichotomous variables for gender (male = 0; female = 1), self-identified minority status (white = 0; African-American or other minority = 1), educational status (fewer than 12 years = 0; 12 or more years of schooling = 1), marital status (not married = 0, married = 1), and employment status (not currently working = 0, currently working = 1). Age was treated as a continuous variable.

The MIMIC model—The MIMIC (multiple indicators, multiple causes) model permits simultaneous estimation of a measurement model (“internal” validity) and the incorporation of external covariates (“external” validity) (Gallo et al. 1994; Hancock 1997; McArdle and Prescott 1992; Muthén 1988b, 1989b). The MIMIC model consists of three components (Fig. 1): (1) a measurement model (a continuous latent variable underlies the dichotomous symptom responses – on the right side of the figure the measurement model relates the observed symptoms to the unobserved latent variable of depression); (2) a regression model (analogous to multiple regression of the continuous outcome variable onto several covariates – on the left side of the figure the regression model relates the latent variable to the covariates); and (3) a “direct effect” estimate (to detect measurement invariance in symptom response associated with membership in a particular group – the path at the top of the figure relates the covariate of interest, gender of the respondent, to a symptom of interest, such as endorsement of “sadness”). The “direct effect” indicates whether “measurement invariance” is present [for a general discussion of measurement invariance, see references (Cunningham 1991; Horn and McArdle 1992; McArdle and Prescott 1992; Muthén 1988b)]. In other words, if the direct effect estimate differs significantly from the null, the measurement of that item differs across groups. For example, the implication of a significant positive direct effect of gender on sadness would be that sadness tends to be more likely to be endorsed by women compared to men, even accounting for the level of depression and the presence of other covariates in the model.

Model estimation procedure and fit—Models were estimated using the Mplus program's limited-information generalized-least squares estimator for dichotomous response (Muthén 1988a, 1989a). The resulting estimates are based on probit regressions on the covariates and manifest indicators of the latent variable (Muthén 1988a). Since even small departures can result in rejection of an adequately fitting model in large samples when χ^2 is used to assess fit, we employed several approaches to assessing model fit; namely, χ^2 (p value greater than 0.05 indicating adequate fit), the descriptive fit value (DFV; value less than 1.5 indicating adequate fit) (Muthén 1989b), and two goodness-of-fit indices (GFI and adjusted GFI; range from 0 to 1 with value greater than 0.90 indicating adequate fit) (Alwin 1988; Bollen 1989).

Results

Study sample

In Baltimore in 1981, 4,238 residents aged 18 years and older were probabilistically designated; 3,481 (82%) completed interviews and were the target sample for follow-up interviews in 1994.

During the follow-up, 848 respondents died, the address of 415 could not be established, 2,218 persons were located, and 1,920 (87%) participated in the interviews. In all, 193 persons were excluded who had incomplete information on symptoms or did not have complete information for all six covariates, leaving 1,727 for this investigation (Table 1). Of the participants, 1,079 (62.5%) were women and 648 (37.5%) were men.

Sociodemographic characteristics of the sample among women and men are shown in Table 2. The age range of the sample was 30–96 years of age (mean age was 52.86 years and standard deviation was 16.04 years).

Prevalence of depressive symptom groups according to gender

Table 3 describes the prevalence of a given symptom group within the last 6 months of interview according to the gender of the participant. The prevalence for each symptom was approximately twice the rate among women compared to men.

Level of the latent trait according to gender and other covariates

Our first step in the analysis was to examine the latent roots or eigenvalues for the sample correlation matrix of the symptoms of depression obtained from exploratory factor analysis with Mplus (equivalent to principal components analysis for dichotomous data). The scree plot (not shown) of the latent roots derived from the covariance matrix of the nine symptom groupings formed from the item parcels corresponding to the diagnostic symptom groups was consistent with a single factor (first eigenvalue 6.04, second eigenvalue 0.65) (Cattell 1985; Edelbrock 1987; Gibbons et al. 1985). The first eigenvalue accounted for 67% of the variance in the symptom data.

The first set of parameter estimates for the MIMIC model concerns the level of the latent trait as captured by the nine depression symptom groupings according to gender and the other covariates. The mean factor level for women compared to men was captured by the regression coefficient for the regression of the latent variable of depression on the covariate for gender, shown in Table 4. As expected, the estimated factor level was higher among women compared to men. Specifically, the adjusted estimated gender difference in factor level was 0.285 [95% confidence interval (CI) 0.152, 0.418] in the model in the right-hand column of Table 4. Differences in factor levels according to other characteristics were also in the directions expected. For example, the coefficients shown in Table 4 indicate that persons with a high-school education or beyond, married persons, and those currently working had lower levels of depression (significantly negative coefficients). Increasing age and self-identified ethnicity as non-white were associated with lower levels of latent trait depression, consistent with similar analyses employing the ECA data (Gallo et al. 1994,1998,1999).

Estimated direct effect of gender on depression symptom groups

We focus first on the estimated direct effect between gender and sadness (top row of Table 5). Although the prevalence of sadness within 6 months of interview by women was about twice that reported by men (Table 3), the confidence interval for the direct effect included the null, suggesting that women were not differentially more likely than men to report sadness [estimated direct coefficient 0.090 (95 % CI −0.106, 0.286)]. Indices for the model containing only the direct effect for gender suggested good fit of the model to the data according to several measures of fit ($\chi^2 = 40.3$, $df = 29$, $p = 0.080$; $DFV = 0.80$; $GFI = 0.99$; adjusted $GFI = 0.98$). After adjustment for covariates, the estimated direct effect was substantially unchanged [0.105 (95% CI −0.113, 0.323)]. Indices for the model that included covariates indicate some degradation of model fit ($\chi^2 = 79.9$, $df = 55$, $p = 0.016$; $DFV = 0.84$; $GFI = 0.99$; adjusted $GFI = 0.97$), but still indicate reasonable fit of the model to the data. All other direct effects to assess

the differential reporting of symptoms by women compared to men were not significantly different from the null.

Discussion

Taking into account differences in the level of depression as indicated by the symptoms reported, as well as other potentially influential covariates, we found no evidence that symptoms of depression tend to be differentially reported at greater rates by women compared to men. This suggests that men and women do not report a different pattern of depressive symptoms that might explain differences in prevalence or incidence rates of major depression in the community. We derived this conclusion by borrowing a model from educational testing to test differential symptom functioning according to gender in the context of the nine symptom groupings that form the diagnostic criteria for major depression from DSM-IV. Specifically, if the model is believed, the higher rates of depression in women are not explained by the tendency of women to report sadness or other depressive symptoms more than men.

Before reflecting on the implications of the findings of this study, we need to call attention to several limitations. With regard to the ascertainment of symptoms, the reports are based on standardized assessments administered by lay interviewers and are subject to imperfect recall and other sources of error. In addition, by limiting analysis to respondents with complete information on covariates in the model, we excluded the most impaired respondents. Our analysis was based on respondents who had survived the 13-year follow-up interval, and these persons might differ from persons who were lost to follow-up in ways that distort the findings. In any case, the direction and extent of any distortion would be difficult to predict, especially in an item-level analysis such as this. There may indeed be differences in the reporting patterns of depressive symptoms among men and women that we did not detect. We realize that we can never “prove” a negative finding. Negative findings require continued analysis of other data sets and other models. Lastly, limitations specific to the MIMIC model for studying differences in symptoms of depression have been discussed at length elsewhere (Gallo et al. 1994; Muthén 1989a).

Limitations notwithstanding, considering a latent trait model of gender differences in depression (that is, a dimensional model) permits the application of analytic methods grounded in item response theory to the issue of the female preponderance of depression. Other methods employed to study group differences in item response are limited in their ability to control for important covariates (e. g., methods based on the Mantel-Haenszel statistic) (Nandakumar et al. 1993; Rogers and Swaminathan 1993; Zwick and Ercikan 1989), to employ a latent variable model with explicit specification of measurement error (e. g., logistic regression) (Camilli and Shepard 1994; Rogers and Swaminathan 1993), or to analyze sparse or skewed symptom data (e. g., multiple-group factor analysis) (Horn and McArdle 1992). Our MIMIC model allowed us to sharpen the focus on differential symptom reporting according to gender.

Given the many previous studies showing an increased prevalence and incidence of depression in women compared to men, it is surprising that few published studies have examined whether symptom patterns differ according to gender. The few other studies that have examined differences between men and women in terms of depression symptom reporting can be divided into two general categories. One set of studies compared total numbers of depressive symptoms between men and women without regard to specific symptoms. These studies have shown that it is not only higher prevalence rates of depressed mood or diminished interest that contribute to the increased prevalence of depression in women, but higher prevalence rates of the total number of depressive symptoms (Chen et al. 2000; Kessler et al. 1993). However, these studies did not examine gender differences for each individual symptom, as we have done here.

A second group of studies has attempted to control for the severity of depression in order to examine whether women with significant clinical depression report more symptoms than do men. When examining individual symptoms during a depressive episode, investigators have found that women are more likely to present with increased appetite and weight gain as well as anxiety and somatic symptoms (Angst and Dobler-Mikola 1984; Frank et al. 1988; Williams et al. 1995; Young et al. 1990). In any case, comparison of depression symptoms according to gender or other personal characteristics will be subject to distortion by factors related to seeking health services. Our strategy allowed us to probe the relationship of gender to particular symptoms while holding constant the influence of other characteristics, such as age and the level of depression among a sample not restricted to persons in the health services.

While our study does not address why women are more prone to major depression, we did not find that women were more likely to report sadness or other symptoms. Many different theories have been proposed to explain the female preponderance of depression. The evidence for all these theories has been thoroughly discussed in several excellent reviews (Blehar and Oren 1995; Nolen-Hoeksema 1987; Nolen-Hoeksema 1990). The biological theories have focused on differences in brain structure and function as well as genetic transmission and reproductive function. The psychosocial theories have focused on the effects of gender-specific socialization, differences in socioeconomic status, differences in role and life stress, and differences in coping styles. There are also artifact theories which examine the possibility that any gender differences in rates of depression are artifacts of differences in help-seeking behavior or differences in symptom-reporting. Our sample was not restricted to persons who sought help in the health services or who met standard criteria for major depression, factors which might distort comparison of symptoms in different groups. Our study supports the notion that gender differences cannot be simply explained as due to differential reporting of depressive symptoms.

Acknowledgments

Data analysis was supported by an American Academy of Family Physicians Advanced Research Training Grant (Dr. Bogner). Data gathering in the Baltimore ECA follow-up (1993–1994) was supported by the National Institute of Mental Health (MH47447).

References

1. Alwin, DF. Structural equation models in research on human development and aging. In: Schaie, KW.; Campbell, RT.; Meredith, W.; Rawlings, SC., editors. *Methodological issues in aging research*. Springer Publishing Company; New York: 1988. p. 71-170.
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, DSM-IV*. 4th. American Psychiatric Association; Washington, DC: 1994.
3. Angst J, Dobler-Mikola A. Do the diagnostic criteria determine the sex ratio in depression? *J Affect Dis* 1984;7:189–198. [PubMed: 6241203]
4. Badawi M, Eaton WW, Myllyluoma J, Weimer LG, Gallo J. Psychopathology and attrition in the Baltimore ECA 15-year follow-up 1981–1996. *Soc Psychiatry and Soc Psychiatry* 1999;34:91–98.
5. Blehar M, Oren D. Women's vulnerability to mood disorders: integrating psychobiology and epidemiology. *Depression* 1995;3:3–12.
6. Bollen, KA. *Structural equations with latent variables*. John Wiley & Sons; New York: 1989.
7. Camilli, G.; Shepard, LA. *Methods for identifying biased test items*. Sage Publishers; Newbury Park, California: 1994.
8. Cattell RB. Factor analysis: an introduction to essentials. *Biometrics* 1965;21:190–215.
9. Chen LS, Eaton WW, Gallo JJ, Nestadt G. Understanding the heterogeneity of depression through the triad of symptoms, course and risk factors: a longitudinal, population-based study. *J Affect Dis* 2000;59:1–11. [PubMed: 10814765]

10. Compas BE, Oppedisano G, Connor JK, Gerhardt CA, Hinden BR, Achenbach TM, Hammen C. Gender differences in depressive symptoms in adolescence: comparison of national samples of clinically referred and nonreferred youths. *J Consult Clin Psychol* 1997;65(4):617–626. [PubMed: 9256563]
11. Cunningham, WR. Issues in factorial invariance. In: Collins, LM.; Horn, JL., editors. *Best Methods for the analysis of change*. American Psychological Association; Washington, DC: 1991. p. 106–113.
12. Duncan-Jones P, Grayson DA, Moran PAP. The utility of latent trait models for psychiatric epidemiology. *Psychol Med* 1986;16:391–405. [PubMed: 3726012]
13. Eaton WW, Anthony JC, Tepper S, Dryman A. Psychopathology and attrition in the Epidemiologic Catchment Area surveys. *Am J Epidemiol* 1992;135:1051–1059. [PubMed: 1595691]
14. Eaton, WW.; Kessler, LG. *Epidemiologic field methods in psychiatry: the NIMH Epidemiologic Catchment Area Program*. Academic Press Inc; New York: 1985.
15. Edelbrock, C. Psychometric research on children and adolescents. In: Last, CG.; Hersen, M., editors. *Issues in diagnostic research*. Plenum Press; New York: 1987. p. 219–240.
16. Fleishman JA, Spector WD, Altman BM. Impact of differential item functioning on age and gender differences in functional disability. *J Gerontol* 2002;57B(5):S275–S284.
17. Frank E, Carpenter LL, Kupfer DJ. Sex differences in recurrent depression: are there any that are significant? *Am J Psychiatry* 1988;145:41–45. [PubMed: 3337291]
18. Gallo JJ, Anthony JC, Muthén BO. Age differences in the symptoms of depression: a latent trait analysis. *J Gerontol: Psychol Sci* 1994;49:P251–264.
19. Gallo JJ, Cooper-Patrick L, Lesikar S. Depressive symptoms of whites and African Americans aged 60 years and older. *J Gerontol: Psychol Sci* 1998;53B:277–286.
20. Gallo JJ, Rabins PV, Anthony JC. Sadness in older persons: 13-year follow-up of a community sample in Baltimore, Maryland. *Psychol Med* 1999;29:341–350. [PubMed: 10218925]
21. Gallo JJ, Royall DR, Anthony JC. Risk factors for the onset of major depression in middle age and late life. *Soc Psychiatry Psychiatr Epidemiol* 1993;28:101–108. [PubMed: 8378804]
22. Gibbons RD, Clark DC, Cavanaugh SV, Davis JM. Applications of modern psychometric theory in psychiatric research. *J Psychiatr Res* 1985;19:43–55. [PubMed: 3989737]
23. Hambleton, RK.; Swaminathan, H.; Rogers, HJ. *Fundamentals of Item Response Theory*. Vol. 2. Sage Publications; Newbury Park, California: 1991.
24. Hancock GR. Structural equation modeling methods for hypothesis testing of latent variable means. *Measurement and Evaluation in Counseling and Development* 1997;30:91–105.
25. Horn JL, McArdle JJ. A practical and theoretical guide to measurement invariance in aging research. *Experiment Aging Res* 1992;18:117–144.
26. Jones RN, Gallo JJ. Education and sex differences in the Mini-Mental State Examination: Effects of differential item functioning. *J Gerontol B Psychol Sci Soc Sci* 2002;57(6):548–558.
27. Jones, RN. Unpublished Sc. D. The Johns Hopkins University School of Hygiene and Public Health; Baltimore, Maryland: 1997. *Validity of the Mini-Mental State Examination*.
28. Kessler RC, McGonagle KA, Swartz M, Blazer DG, Nelson CB. Sex and depression in the National Comorbidity Survey: I. Lifetime prevalence, chronicity, and recurrence. *J Affect Dis* 1993;29:85–96. [PubMed: 8300981]
29. Kornstein SG, Schatzberg AF, Thase ME, Yonkers KA, McCullough JP, Keitner GI, Gelenberg AJ, Ryan CE, Hess AL, Harrison W, Davis SM, Keller MB. Gender differences in chronic major and double depression. *J Affect Dis* 2000;60:1–11. [PubMed: 10940442]
30. McArdle JJ, Prescott CA. Age-based construct validation using structural equation modeling. *Experiment Aging Res* 1992;18:87–115.
31. Muthén, BO. *LISCOMP: analysis of linear structural equations with a comprehensive measurement model*. Scientific Software; Mooresville, Indiana: 1988a.
32. Muthén, BO. Some uses of structural equation modeling in validity studies: extending IRT to external variables. In: Wainer, H.; Braun, H., editors. *Test Validity*. Lawrence Erlbaum; Hillsdale, NJ: 1988b. p. 124–144.
33. Muthén BO. Dichotomous factor analysis of symptom data. *Sociol Methods Res* 1989a;18:19–65.

34. Muthén BO. Latent variable modeling in heterogeneous populations. *Psychometrika* 1989b;54:557–585.
35. Nandakumar R, Glutting JJ, Oakland T. Mantel-Haenszel methodology for detecting item bias: an introduction and example using the guide to the assessment of test session behavior. *J Psychoeducational Assess* 1993;11:108–119.
36. Nolen-Hoeksema S. Sex differences in unipolar depression: evidence and theory. *Psychol Bull* 1987;101:259–282. [PubMed: 3562707]
37. Nolen-Hoeksema, S. Sex differences in depression. Stanford University Press; Stanford, CA: 1990.
38. Petty F, Nasrallah HA. Secondary depression in alcoholism: implications for future research. *Comprehens Psychiatry* 1981;22:587–595.
39. Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. *Arch Gen Psychiatry* 1981;38:381–389. [PubMed: 6260053]
40. Robins, LN.; Regier, DA. Psychiatric disorders in America: the Epidemiologic Catchment Area Study. Free Press; New York: 1991.
41. Rogers HJ, Swaminathan H. A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *App Psychol Measurement* 1993;17:105–116.
42. Schaeffer, NC. An application of item response theory to the measurement of depression. In: Clogg, CC., editor. *Sociological Methodology*. Vol. 18. Jossey-Bass; San Francisco, CA: 1988. p. 271-307.
43. Suh T, Gallo JJ. Symptom profiles of depression among general medical service users compared with specialty mental health service users. *Psychol Med* 1997;27:1051–1063. [PubMed: 9300511]
44. Van der Linden, WJ.; Hambleton, RK., editors. *Handbook of Item Response Theory*. Springer-Verlag; New York: 1997.
45. Von Korff M, Anthony JC. The NIMH Diagnostic Interview Schedule modified to record current mental status. *J Affect Dis* 1982;4:365–371. [PubMed: 6298294]
46. Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu HG, Joyce PR, Karam EG, Lee CK, Lellouch J, Lepine JP, Newman SC, Rubio-Stipec M, Wells JE, Wickramaratne PJ, Wittchen HU, Yeh EK. Cross-national epidemiology of major depression and bipolar disorder. *JAMA* 1996;276:293–299. [PubMed: 8656541]
47. Weissman MM, Klerman GL. Sex differences in the epidemiology of depression. *Arch Gen Psychiatry* 1977;34:98–111. [PubMed: 319772]
48. Weissman MM, Klerman JK. Depression: current understanding and changing trends. *Ann Rev Public Health* 1992;13:319–339. [PubMed: 1599592]
49. West, SG.; Finch, JF.; Curran, PJ. Structural equation models with nonnormal variables: problems and remedies. In: Hoyle, RH., editor. *Structural equation modeling: concepts, issues, and applications*. Sage; Thousand Oaks, California: 1995. p. 56-75.
50. Williams JB, Spitzer RL, Linzer M, Kroenke K, Hahn SR, deGruy FV, Lazev A. Gender differences in depression in primary care. *Am J Obstetr Gynecol* 1995;173:654–659.
51. Wittchen HU. Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): a critical review. *J Psychiatr Res* 1994;28:57–84. [PubMed: 8064641]
52. Young MA, Scheftner WA, Fawcett J, Klerman GL. Gender differences in the clinical features of unipolar depressive disorder. *J Nerv Ment Dis* 1990;178:200–203. [PubMed: 2307973]
53. Zwick R, Ercikan K. Analysis of differential item functioning in the NAEP history assessment. *J Educational Measurement* 1989;26:55–66.

Fig. 1.
The multiple indicators multiple causes (MIMIC) model

Table 1

Totals and percent available for analysis in models including all covariates

	Total from Baltimore ECA follow-up study sample	% available for the present analysis
Age 65 years and older	369	99.7
Women	1,215	89.0
Self-identified minority	706	92.5
Education less than high school	863	84.1
Married	865	98.6
Currently working	905	97.7
Total	1,920	90.0

Table 2

Characteristics of the sample for analysis in models including all covariates. Percents represent gender-specific proportions of given characteristics

	Women		Men	
Total available	1,079		648	
	n	%	n	%
Age 65 years and older	322	29.8	154	23.8
Self-identified minority	449	41.6	204	31.5
Education less than high school	473	43.8	253	39.0
Married	452	41.9	401	61.9
Currently working	486	45.0	398	61.4

Table 3

Prevalence of given symptom group within the last 6 months of interview, according to gender of the respondent. Numbers in parentheses represent column percents. Data gathered from the Baltimore, Maryland Epidemiologic Catchment Area Program follow-up, 1994

	Women (n = 1,079)	Men (n = 648)
Sadness	87 (8.1%)	23 (3.5%)
Appetite disturbance	95 (8.8%)	41 (6.3%)
Sleep disturbance	133 (12.3%)	50 (7.7%)
Fatigue	91 (8.4%)	30 (4.6%)
Psychomotor change	54 (5.0%)	19 (2.9%)
Loss of interest	155 (14.4%)	43 (6.6%)
Worthlessness	59 (5.5%)	19 (2.9%)
Trouble concentrating	60 (5.6%)	15 (2.3%)
Thoughts of death or suicide	84 (7.8%)	27 (4.2%)

Table 4

Parameter estimates for MIMIC models. Data gathered from the Baltimore, Maryland Epidemiologic Catchment Area Program follow-up, 1994. 95 % confidence intervals are given in brackets

Estimated differences in the mean level of depression in the regression component of the MIMIC model		
Female (reference group: males)	0.331 [*] (0.200, 0.462)	0.285 [*] (0.152, 0.418)
Age 65 years and older (reference group: age < 65 years)		-0.014 [*] (-0.018, -0.010)
African-American or other minority (reference group: white)		-0.028 (-0.159, 0.103)
Education high school and beyond (reference group: education less than high school)		-0.186 [*] (-0.325, -0.047)
Married (reference group: not married)		-0.192 [*] (-0.323, -0.061)
Currently working (reference group: not working)		-0.318 [*] (-0.473, -0.163)

*
p < 0.05

Table 5

Estimated direct effect of gender on endorsement of symptom criteria in MIMIC model. Model 1 includes only sex as a covariate. Estimates in model 2 are adjusted for the other potentially influential covariates listed in Table 3. Data gathered from the Baltimore, Maryland Epidemiologic Catchment Area Program follow-up, 1994. 95 % confidence intervals are given in brackets

	Model 1	Model 2
Sadness	0.090 (-0.106, 0.286)	0.105 (-0.113, 0.323)
Appetite disturbance	-0.108 (-0.288, 0.072)	-0.100 (-0.284, 0.084)
Sleep disturbance	-0.054 (-0.221, 0.113)	-0.069 (-0.240, 0.102)
Fatigue	-0.035 (-0.206, 0.136)	-0.049 (-0.229, 0.131)
Psychomotor change	-0.106 (-0.324, 0.112)	-0.101 (-0.330, 0.128)
Loss of interest	0.149 (-0.014, 0.312)	0.171 (-0.001, 0.343)
Worthlessness	-0.056 (-0.254, 0.142)	-0.032 (-0.244, 0.180)
Trouble concentrating	0.066 (-0.150, 0.282)	0.079 (-0.162, 0.320)
Thoughts of death or suicide	0.036 (-0.168, 0.240)	-0.001 (-0.217, 0.215)