Cross-national associations between gender and mental disorders in the WHO World Mental Health Surveys

Soraya Seedat, PhD[1], **Kate Margaret Scott, PhD**[2], **Matthias C. Angermeyer, PhD**[3], **Patricia Berglund, MBA**[4], **Evelyn J. Bromet, Ph.D.**[5], **Traolach S. Brugha, MD (NUI), FRCPsych**[6], **Koen Demyttenaere, MD, PhD**[7], **Giovanni de Girolamo, MD**[8], **Josep Maria Haro, MD, MPH, PhD**[9], **Robert Jin, MA**[10], **Elie G. Karam, MD**[11], **Viviane Kovess-Masfety, MD, PhD**[12], **Daphna Levinson, PhD**[13], **Maria Elena Medina Mora, PhD**[14], **Yutaka Ono, MD, PhD**[15], **Johan Ormel, PhD**[16], **Beth-Ellen Pennell, MA**[4], **Jose Posada-Villa, MD**[17], **Nancy A. Sampson, BA**[10], **David Williams, PhD, MPH**[18], and **Ronald C. Kessler, PhD**[10]

[1] MRC Research Unit on Anxiety and Stress Disorder, Cape Town, South Africa [2] Department of Psychological Medicine, Wellington School of Medicine, Otago University, Wellington, New Zealand [3] Center for Public Mental Health, Austria [4] Institute for Social Research, University of Michigan, Ann Arbor, MI [5] SUNY Stony Brook, Stony Brook, NY [6] Department of Health Sciences, University of Leicester, United Kingdom [7] University Hospital, Gasthuisberg, Leuven, Belgium [8] Health Care Research Agency, Bologna, Italy [9] Sant Joan de Deu-SSM, CIBERSAM, Instituto de Salud Carlos III, Barcelona, Spain [10] Department of Health Care Policy, Harvard Medical School, Boston, MA [11] Institute for Development, Research, Advocacy, and Applied Care (IDRAAC) St., George Hospital University Medical Center, Lebanon [12] University of Paris Descartes, EA 4069; MGEN Foundation for Public Health, Paris, France [13] Ministry of Health, Mental Health Services, Israel [14] Department of Epidemiology, National Institute of Psychiatry, Mexico City, Mexico [15] Keio University School of Medicine, Tokyo, Japan [16] University Medical Center Groningen, University of Groningen, Groningen, The Netherlands [17] Colegio Mayor de Cundinamarca University, Bogata, Colombia [18] Harvard University School of Public Health, Boston, MA

Abstract

Context—Gender differences in mental disorders, including more anxiety-mood disorders among women and more externalizing disorders among men, are found consistently in epidemiological surveys. The "gender roles" hypothesis suggests that these differences should narrow as the roles of women and men become more equal.

Objective—To study time-space (i.e., cohort-country) variation in gender differences in lifetime DSM-IV mental disorders across cohorts in 15 countries in the WHO World Mental Health (WMH) Survey Initiative and determine if this variation is significantly related to time-space variation in female gender role traditionality (GRT) as measured by aggregate patterns of female education, employment, marital timing, and use of birth control.

Design/Setting and Participants—Face-to face household surveys of 72,933 community-dwelling adults in Africa, the Americas, Asia, Europe, the Middle East, and the Pacific.

Main Outcomes—The WHO Composite International Diagnostic Interview (CIDI) assessed lifetime prevalence and age-of-onset of 18 DSM-IV anxiety, mood, externalizing, and substance disorders. Survival analyses estimated time-space variation in Female:Male (F:M) odds-ratios (ORs) of these disorders across cohorts defined by age ranges 18–34, 35–49, 50–64, and 65+. Structural equation analysis examined predictive effects of variation in GRT on these ORs.

Results—Women had more anxiety-mood disorders than men and men more externalizing-substance disorders than women in all cohorts and countries. Although gender differences were generally consistent across cohorts, significant narrowing was found in recent cohorts for major depressive disorder (MDD) and substance disorders. This narrowing was significantly related to temporal (MDD) and spatial (substance disorders) variation in GRT.

Conclusion—While gender differences in most lifetime mental disorders were fairly stable over the time-space units studied, substantial inter-cohort narrowing of differences in major depression was found related to changes in the traditionality of female gender roles. Further research is needed to understand why this temporal narrowing was confined to major depression.

Epidemiological surveys have consistently documented significantly higher rates of anxiety and mood disorders among women than men[1, 2] and significantly higher rates of externalizing and substance use disorders among men than women.[3–5] Although a number of biological, psychosocial, and biopsychosocial hypotheses have been proposed to account for these patterns,[6–8] evidence that gender differences in depression[9, 10] and substance use[11–13] have narrowed in a number of countries has led to a special interest in the "gender roles" hypothesis. The latter asserts that gender differences in the prevalence of mental disorders are due to differences in the typical stressors, coping resources, and opportunity structures for expressing psychological distress made available differentially to women and men in different countries at different points in history.[14, 15] Consistent with this hypothesis, evidence of decreasing gender differences in depression and substance use has been found largely in countries where the roles of women have improved in terms of opportunities for employment, access to birth control, and other indicators of increasing gender role equality, while trend studies in countries where gender roles have been more static[11, 16] or over periods of historical time when gender role changes have been small[17] have failed to document a reduction in gender differences in depression or substance use.

Most research aimed at investigating the gender roles hypothesis has focused on individual-level variation in roles in a single country at a single point in time.[18–20] This approach is limited in three ways. First, selection bias into roles due to pre-existing mental illness (e.g., women with agoraphobia having a higher probability than other women of becoming homemakers rather than seeking employment outside the home) confounds attempts to evaluate the causal effects of gender roles. Second, gender differences are largely confined to differences

in lifetime risk, with much less evidence for gender differences in recent prevalence among lifetime cases.[21] This means that investigation of the determinants of gender difference should focus on lifetime first onset rather than on the recent prevalence that has been the focus of most studies. Third, as the gender roles hypothesis is a hypothesis about the effects of social context, a rigorous test of the hypothesis requires an analysis of societal-level time-space variation rather than analysis of the individual-level variation that has been the focus of most studies.

A small number of cross-national comparative studies have examined spatial variation in gender differences in depression[22] and alcohol abuse[13] at a point in time or, more rarely, at two points in time.[11] Although these studies raised the possibility that gender roles might be associated with variation in the magnitude of gender differences in these outcomes, they were unable to test this hypothesis due to the small number of cross-sectional country-level observations included in the analyses. The current report provides a more direct test of the gender roles hypothesis by analyzing community epidemiological data collected from respondents surveyed in 15 countries as part of the World Health Organization (WHO) World Mental Health (WMH) Survey Initiative.[21] Previous cross-national comparisons of gender differences in mental illness focused on cross-sectional differences. We, in comparison, use retrospective reports obtained in the WMH surveys about lifetime occurrence and age-of-onset of mental disorders in different birth cohorts to study time-space variation in lifetime risk. Specifically, we examine both variation across cohorts within a single country (i.e., temporal variation) and variation across countries within a single cohort (i.e., special variation) in lifetime risk of mental disorders as a function of time-space variation in the traditionality of gender roles. Lifetime risk is the focus rather than recent prevalence even though accuracy of reporting is doubtlessly better for recent episodes than lifetime occurrence in order to address the fact that gender differences in lifetime risk are much more robust than gender differences in current prevalence among lifetime cases

METHODS

Sample

WMH surveys were carried out in samples of adults (ages 18+) in five countries classified by the World Bank[23] as developing (Colombia, Lebanon, Mexico, South Africa, Ukraine) and ten classified as developed (Belgium, France, Germany, Israel, Italy, Japan, Netherlands, New Zealand, Spain, and United States of America). The total sample size was 72,933. Individual country samples ranged from 2,372 (Netherlands) to 12,790 (New Zealand). (Table 1) The weighted average response rate was 71.2%. Country-specific response rates ranged from 45.9% (France) to 87.7% (Colombia). All surveys were based on probability household samples either regionally representative (Colombia, Japan, and Mexico) or nationally representative (other countries). Survey sample characteristics are described in more detail elsewhere.[24]

All interviews were conducted face-to-face by trained lay interviewers. Each interview had two parts. All respondents completed Part I, which contained assessments of core mental disorders, while all Part I respondents who met criteria for any core disorder plus a probability sub-sample of approximately 25% of other Part I respondents were administered Part II. Part II assessed correlates, service use, and disorders of secondary interest. The Part II data, used in the current report, were weighted to adjust for over-sampling Part I respondents with mental disorders, for differential probabilities of selection within households (due to only one household member, and in some cases two, being surveyed in each household no matter the number of adults residing in the household), and to match sample distributions to population socio-demographic distributions. Standardized interviewer training procedures, translation and back-translation procedures, and quality control procedures were applied across all WMH countries to ensure comparability. Informed consent was obtained in all countries. These procedures are described in more detail elsewhere.[24, 25]

DSM-IV disorders

Mental disorders were assessed with Version 3.0 of the WHO Composite International Diagnostic Interview (CIDI),[26] a fully-structured diagnostic interview. Translation, back-translation, and harmonization of the interview in local languages with the original English language version of CIDI 3.0 were carried out in each WMH country using WHO guidelines. [27] Disorders were assessed using the definitions of the American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV).[28] Disorders assessed included mood disorders (major depressive disorder, dysthymic disorder, bipolar disorder), anxiety disorders (panic disorder, generalized anxiety disorder, agoraphobia without panic disorder, social phobia, specific phobia, separation anxiety disorder, post-traumatic stress disorder), externalizing disorders (attention-deficit/hyperactivity disorder, conduct disorder, intermittent explosive disorder, oppositional-defiant disorder), and substance disorders (alcohol and illicit drug abuse with or without dependence). DSM-IV organic exclusion rules were used to make diagnoses.

Methodological evidence collected in clinical reappraisal studies show that diagnoses of anxiety, mood, and substance disorders based on CIDI 3.0 have generally good concordance (25th–75th percentiles of area under the receiver operating characteristic curve equal to .71–.81) with diagnoses based on blinded clinical reappraisal interviews.[29] No evaluations were made of test-retest reliability. The evidence regarding good concordance with clinical diagnoses is based on surveys carried out in only a small number of countries. It is not clear that the translations of the instrument in all countries yield data that would have the same good concordance with blinded clinical reappraisal interviews. In addition, the externalizing disorder diagnoses were not validated in the CIDI 3.0 clinical reappraisal studies because the clinical interview used as the gold standard in these studies did not assess externalizing disorders. However, a subsequent independent clinical calibration study documented good concordance between diagnoses of adult ADHD based on CIDI 3.0 and those based on blinded clinical reappraisal interviews.[30] Another problem exists with the diagnoses of substance dependence, which were assessed only among respondents who had a history of abuse. This means that cases of dependence without abuse are excluded. However, empirical studies in the US have shown that the number of cases of dependence without a history of abuse is small and that their exclusion does not have a substantively meaningful effect on the estimated associations of predictors with the outcomes.[31–33] Nonetheless, because of this exclusion we focus here on abuse rather than dependence. Retrospective age-of-onset (AOO) reports were obtained in the CIDI using a series of questions designed to avoid the implausible response patterns obtained in response to a simple question asking for recall of age of first episode of a focal disorder.[34]

Gender role traditionality

Each respondent was classified as being in one of four birth cohorts defined by age at interview (18–34, 35–49, 50–64, 65+) to distinguish broad life course stages (early adulthood, early midlife, late midlife, and old age). Four within-cohort indicators of female gender role traditionality were calculated in each of the 58 resulting time-space subsamples (i.e., four cohorts in each of 15 countries, minus the two oldest cohorts in Colombia and Mexico due to an upper age limit of 64 in those two surveys). The four were as follows: (i) the ratio of the proportion of women to men in the cohort who had labor force experience before age 35 (extrapolated using survival analysis in the 18–34 cohort and calculated directly in the older cohorts); (ii) the ratio of the proportion of women to men in the cohort who achieved the median level of education found among workers in the upper quartile of the income distribution in the cohort; (iii) the ratio of the median ages of marriage of women versus men in the cohort; and (iv) the proportion of women in the cohort who used birth control pills or other medical forms of contraception before age 25 (restricted to women ages 25–34 in the 18–34 cohort).

We make no claim that these indicators form an exhaustive set of the defining characteristics of gender role traditionality or that the cut-points used to construct the measures (e.g., labor force participation by age 34 rather than by some other age that we might have selected) are optimal. Rather, the indicators were constructed from the WMH survey data on an ad hoc basis to operationalize aspects of gender roles that we considered important based on our reading of the demographic literature on gender roles.[35–37]

Confirmatory factor analysis carried out at the level of the time-space unit (n = 58) showed that our initial thinking in selecting the four indicators was correct in the sense that a strong single-factor structure was found among these four indicators, with factor loadings in the range .59–.91. (Detailed results showing the values of each GRT indicator for each time-space unit, the correlation matrix among the indicators, and factor loadings are available on request.) This finding confirmed that the indicators are, in fact, strongly related and can be used to construct a composite measure that we interpret as a measure of gender role traditionality (GRT). This is the key predictor variable in the analysis described below.

Rather than use the ad hoc GRT measure described above, it would have been preferable to obtain objective administrative data on country-level trends in indicators of GRT. Our attempts to obtain such data, though, were unsuccessful because of sparse historical data on these indicators in most countries. The Global Economic Forum (GEF) collected country-level data of this sort to assess the social-economic-political positions of women in 58 countries in the year 2000,[38] but they, like us, were unable to obtain retrospective trend data. The goal of the GEF undertaking was to create a baseline measure that could be used to track the UN Millennium Development Goals of gender equality in social, economic, and political functioning (www.un.org/milleniumgoals). The GEF report constructed a country-level *Gender Empowerment Measure* (GEM) for this purpose that summarized objective data on the economic opportunity and participation of women in each country along with data on female political empowerment, educational attainment, life expectancy, and access to health care (legal birth control and legal abortion). The GEM measure could be developed for only 58 countries because of missing data in the other countries of the world. GEM scores are unavailable for two WMH countries (Lebanon and Ukraine). As the GEM measure was developed only for the year 2000, we could not use it to study within-country changes in gender equality over time. However, we were able to compare scores on the composite WMH GRT measure with GEM scores for the 13 WMH countries where GEM scores were available in an effort for validate our survey-based measure against the gold standard GEM measure. The Pearson correlation between the two measures was found to be .78. This high correlation strongly suggests that our GRT measure validly assesses the traditionality of the roles of women in the WMH countries.

Statistical Analyses

Gender differences in lifetime risk of each disorder were examined using discrete-time survival analysis with person-year as the unit of analysis.[39] This is a method that takes into consideration age of onset of the disorder in examining predictors, making it possible to study the predictors of lifetime occurrence of the disorder among respondents who vary in age. Each year in the life of each respondent up to and including the age of onset of the focal disorder (or, in the case of respondents who never had the disorder, up to their age at interview) was treated as a separate observational record in this analysis, with the year of first onset coded 1 on a dichotomous outcome variable and earlier years coded 0. Years after first onset were excluded from the data file. Logistic regression analysis was used to analyze these data, with gender (coded 1 for women, 0 for men), cohort (coded into the four categories noted above), and person-year (age at the time of the person-year observational record) included as predictors of first onset of the disorder. The logistic regression coefficients and their standard errors were exponentiated to

create odds-ratios (ORs) and 95% confidence intervals for ease of interpretation. Female:Male (F:M) ORs are the main focus of attention.

A separate model was estimated for each DSM-IV/CIDI disorder separately in each country. These models were also estimated in a data file that combined observations across all countries. The cross-national models included 14 dummy predictor variables to distinguish among the 15 countries in addition to the other predictors. The basic models were elaborated to consider possible non-linear effects of cohort and person-year (using polynomials and dummy variables to define ranges on these continuous variables) and to assess whether the gender difference in lifetime risks of the disorders varied by cohort, life course phase, or country. Gender differences were for the most part consistent across the life course, so these results are not reported here but are available on request. The models were then estimated a final time in the sub-sample of person-years in the age range 1–34 (i.e., up to the oldest age in the youngest grouped cohort sub-sample) so as to remove the association between cohort and age in the person-year data file.

In cases where the survival analysis documented significant time-space variation in the gender difference for a particular outcome, structural equation models (SEMs) using the 58 time-space sub-samples as the unit of analysis estimated the extent to which a latent measure of gender role traditionality (defined in terms of the four indicators described above) could account for this variation. The best-fitting model was determined as the one with the lowest value on the Bayesian Information Criterion (BIC),[40] a standard measure of model fit. SEMs are regression models that estimate coefficients simultaneously across a series of equations, some of which can include presumed latent (i.e., not directly measured) variables that are assumed to have a pre-specified relationship to measured variables, in an effort to maximize the fit between predicted and observed matrices of covariation among the observed variables. In our case, the SEMs assumed that time-space variation in a latent measure of gender role traditionality, which was indicated by the four observed measures described above, predicted time-space variation in the F:M ORs of disorders that were found to have significant time-space variation.

Survival coefficients and their standard errors were estimated using the Taylor series linearization method[41] implemented in the SUDAAN software system.[42] Multivariate significance tests of the significance of interactions involving gender with person-year, cohort, and country were made with Wald χ^2 tests using Taylor series design-based coefficient variance-covariance matrices. In the case of cross-national models, a single variable was created that assigned a unique value to each sampling stratum in each country, while a second variable was created that distinguished sampling-error calculation units (SECUs) within each stratum. These two variables were used as the input to SUDAAN to calculate design-based estimates. SEMs were calculated using the Mplus software system www.statmodel.com).[43] Significance tests of regression coefficients in the SEMs were estimated using the standard errors generated by Mplus, which assumed that the 58 time-space observations represent a simple random sample from a larger universe of such units. All significance tests were evaluated at the .05 level with two-sided tests.

RESULTS

Gender differences in lifetime risk

Results are highly consistent across countries in showing that women have a significantly higher lifetime risk than men of most mood disorders (major depressive disorder and dysthymic disorder) and all anxiety disorders (Table 2). The pooled F:M ORs for these disorders are all statistically significant and in the range 1.3–2.6. Within-country ORs for these disorders are also consistently greater than 1.0. The one exception to this general pattern is bipolar disorder, where the pooled OR is not statistically significant (0.9). Results are the opposite for most

externalizing disorders (ADHD, conduct disorder, intermittent explosive disorder) and all substance disorders. The pooled F:M ORs for these disorders are all statistically significant and less than 1.0 (in the range 0.3–0.7), indicating significantly higher risk among men than women. Within-country ORs for these disorders are also consistently (100%) less than 1.0. Despite their consistent direction, the magnitudes of the ORs vary significantly across countries for many of the disorders studied.

Inter-cohort variation

Analysis of interactions between gender and cohort show that gender differences in lifetime risk for most disorders do not differ significantly across cohorts. (Table 3) However, there are three notable exceptions. The first involves major depressive disorder (MDD), where the pooled OR for the gender-by-cohort interaction across countries is 0.88 (0.82–0.95). This means that the higher odds among women than men are less pronounced in more recent than older cohorts. This general pattern is found in 11 of the 15 countries. The second involves intermittent explosive disorder (IED), where the pooled gender-by-cohort OR across countries is 1.26 (1.07–1.48). This means that the higher odds among men than women are less pronounced in more recent than older cohorts. This general pattern is found in 5 of the 6 countries that assessed IED. The third involves substance disorders, largely driven by alcohol disorders. The pooled gender-by-cohort OR across countries in predicting any substance disorder is 1.45 (1.27–1.66). It is important to remember that "any" substance disorder is equivalent to either alcohol or drug abuse because, as noted above in the section on measures, dependence was assessed only among respondents with a history of abuse. This means that the higher odds of a lifetime substance abuse among men than women are less pronounced in more recent than older cohorts. This general pattern is found in 12 of the 14 countries where substance abuse was assessed.

Time-space variation in gender role traditionality and F:M ORs

The remaining analyses focused on MDD and substance disorders, the disorders associated with significant inter-cohort variation in the F:M ORs across the majority of countries. (IED was excluded because it was assessed in only six WMH countries, yielding too few time-space units for stable analysis). Before turning to the results, it is instructive to examine the raw data on time-space variation in the composite measure of GRT created by averaging standardized scores on the four indicators. (Table 4) A generally monotonic decrease in traditionality can be seen across successively more recent cohorts in each country. All but two countries (New Zealand and Ukraine) were above the mean in GRT at the time respondents in the oldest cohorts were in early adulthood. All but two countries were below the mean and the other two (Italy and Lebanon) very close to the mean GRT, in comparison, by the time respondents in the most recent cohorts entered early adulthood. Lebanon had by far the highest traditionality score in each cohort, but the other countries with high GRT scores in the oldest cohorts were all developed countries either in Southern Europe (Italy and Spain) or Asia (Japan). Three of these four countries (Japan, Lebanon, and Spain) had the most dramatic decreases in GRT over time, along with Belgium and the Netherlands.

A number of structural equation models were fit to examine the associations of this time-space variation in GRT with the F:M ORs in MDD and substance disorder. (Detailed results are available on request.) The final model (Figure 1) defines GRT as a standardized (to a mean of 0 and variance of 1) latent variable indicated by our four observed measures, where GRT is a predictor of the F:M ORs of MDD and any substance disorder. The four measured indicators are assumed to have effects on these outcomes only through GRT. A high GRT score is defined as *high* gender role traditionality, which is significantly associated with an *increase* of .38 standard deviations (SD) in the F:M OR for MDD (i.e., the female excess in MDD decreases as female gender roles become less traditional) and with a *decrease* of .46 SD in the F:M OR

for any substance disorder (i.e., women begin to "catch up" to men in their rates of substance disorder as female gender roles become less traditional).

As all coefficients in the model are standardized (i.e., the measures are transformed to have a mean of 0 and a variance of 1), it is necessary to consider the substantive meaning of a standard deviation on each measure in order to put the results into meaningful terms. Beginning with the GRT indicators, a one SD decrease in GRT would be equivalent to changing the F:M ratio of labor force participation from the sample-wide mean of .85 (i.e., women about 15% less likely than men to be in the labor force) to 1.0 (i.e., women and men equally likely to be in the labor force), changing the F:M ratio of high educational attainment from the mean of .84 (i.e., women 16% less likely than men to obtain high education) to 1.03 (i.e., rough gender equality), changing the older age at marriage of men than women from the mean of 3.2 years to 2.2 years, and changing the proportion of young women using birth control from the mean of 37.5% to 69.4%.

According to the model, a decrease of one standard deviation in the GRT indicators is associated with a decrease in the F:M OR of MDD of .38 SD and an increase in the F:M OR of any substance disorder of .46 SD. We need to know the means and SDs of the outcomes in the un-weighted sample of 58 time-space units to make substance sense of these effect size estimates. These values are 2.61 (Mean) and 1.81 (SD) for MDD and 0.17 (Mean) and 0.14 (SD) for any substance disorder. Therefore, changes of one SD in the GRT indicators are associated with a reduction in the F:M OR for MDD from the mean of 2.6 to 1.9 (a reduction of nearly 45% in the elevated F:M OR) and with an increase in the F:M OR of any substance disorder from the mean of 0.17 to 0.25 (a reduction of nearly 30% in the elevated M:F OR).

Although variations due to time and space were combined in the structural equation analysis, it is possible to separate the two components by introducing dummy variable controls in the model either for cohort (time) or for country (space). (Table 5) When this is done, we see that the association between GRT and the F:M OR for MDD is entirely due to between-cohort variation within countries, while the association between GRT and the F:M OR for any substance disorder is largely due to within-cohort variation across countries. It is consequently only the association involving MDD that involves inter-cohort changes within countries. We also investigated the possibility that the results could be sensitive to extreme values in a small number of countries by replicating the results in Table 5 15 times, each time deleting the data from one country. (Detailed results are available on request.) The only case in which a meaningful change in the model coefficients occurred was when we deleted Lebanon, but even in this case the coefficients remained statistically significant and strong in substantive terms. This result demonstrates that the overall study results are not highly sensitive to individual outlier countries.

COMMENT

Several methodological limitations need to be noted in interpreting the WMH results. First, the response rates were lower in developed than developing countries and might have been related to gender role traditionality, possibly introducing bias into results. We weighted the data in each country for differential non-response by census socio-demographic variables, but there is no guarantee that this corrected for biases introduced by incomplete response. Second, the surveys also differed across countries in other ways, such as the language in which the survey was administered and the extent to which the country had a tradition of independent public opinion research that allowed respondents to see the survey as a normative undertaking, that might have affected results. Third, diagnoses were based on fully-structured interviews administered by lay interviewers rather than on clinician-administered semi-structured interviews. This limitation is somewhat reduced by the fact that WMH clinical reappraisal

studies documented generally good concordance between the diagnoses based on the CIDI interviews and diagnoses based on blinded semi-structured clinical reappraisal interviews.[29] As noted in the section on measures, though, the diagnoses of externalizing disorders were not validated and might be less accurate than the diagnoses of other disorders. The substance dependence diagnoses had the additional problem of excluding cases of dependence without a history of abuse. The results concerning the broadly-defined measures of substance disorders should consequently be interpreted as applying to abuse rather than to dependence. As noted in the section on measures, though, empirical studies have shown that the number of cases of substance dependence without a history of abuse is small in the US and that their exclusion does not meaningfully affect the size of the coefficients between predictors and measures of substance disorders. Fourth, lifetime prevalence and age of onset were assessed with retrospective reports, which could be systematically biased.[44] We used an innovative probing method designed to minimize recall bias,[34] but it might still be that bias was introduced by age-related gender differences either in memory failure, mental health awareness, or willingness to admit emotional problems to an interviewer. Fifth, the analysis of gender differences in lifetime risk across cohorts could be biased if the increasing attrition with age in older cohorts is differentially related to history of mental disorders among women versus men. Sixth, the indicators of female gender role traditionality were few in number and might not have captured all the dimensions of female gender roles that are important for explaining secular trends in the F:M ORs. The indicators we used, furthermore, might be related to constructs other than gender role traditionality. These concerns are reduced, though, by the fact that our composite GRT measure correlates very strongly with an independent measure of Gender Empowerment based on objective administrative data assembled by the Global Economic Forum.

Within the context of the above limitations, the current report is the first to present the results of a quantitative examination of time-space variation in the association between female gender role traditionality and gender differences in mental illness. We found that the frequently observed gender differences in anxiety, mood, externalizing, and substance disorders have remained relatively stable over the more than half century separating respondents in the youngest and oldest WMH cohorts despite the fact that unprecedented changes occurred over this time period in female gender roles. Furthermore, we found that aggregate F:M ORs are relatively consistent across countries despite substantial between-country variation in female gender role traditionality. These patterns argue against the claim that changes in gender roles play an important role in bringing about reductions in gender differences in the lifetime risk of most mental disorders.

The only notable exceptions to this general pattern concern major depression, intermittent explosive disorder, and substance disorders, where gender differences were found to be significantly smaller in more recent than earlier cohorts. As noted in the introduction, evidence consistent with this narrowing has been found in several within-country studies of gender differences in MDD,[9, 10, 45] substance use,[11] and substance disorders,[46–48] while other studies have documented cross-national variation in gender differences at a point in time.[13, 22] No previous study, though, combined the two types of comparisons to study time-space variation while linking independent measures of gender role traditionality with data on variation in gender differences over time or space.

In the case of MDD, the gender roles hypothesis would interpret our findings as meaning that increases in female opportunities in the domains of employment, birth control, and other indicators of increasing gender role equality promote improvements in female mental health by reducing exposure to stressors that can led to depression and by increasing access to effective stress-buffering resources.[15, 49] It is important to acknowledge, though, that we did not directly evaluate the validity of this hypothesis. We documented that gender differences in MDD onset

risk are significantly narrower at times and places when the roles of women are more equal to those of men, but we did not measure time-space variation in stress exposure or stress reactivity to see if the latter mediate the predictive effects of gender role traditionality. These more fine-grained analyses go beyond the limits of the WMH data, but should be the subject of future studies.

In the case of substance disorders, the gender roles hypothesis takes a somewhat different form in arguing that opportunities for female substance use and attitudes about the appropriateness of female substance use both change as female roles become more similar to male roles, resulting in an increase in female substance use.[14, 50, 51] Consistent with this hypothesis, cross-national research has documented that current female drinking behavior is more similar to male drinking behavior in countries where female and male gender roles are more equal,[52, 53] although there results have not been entirely consistent.[54] These previous studies did not examine gender differences in lifetime risk of substance disorders, though, which means that our results can be seen as building on the findings of these earlier studies.

It is unclear why the WMH findings showed much stronger evidence for temporal than spatial predictive associations for MDD and for spatial than temporal associations for substance disorders. It is also unclear why the narrowing of gender differences in recent cohorts was confined to MDD, IED, and substance disorders. It is conceivable that the much earlier ages of onset of most other disorders, especially the anxiety disorders and externalizing disorders, than MDD or substance disorders makes the former disorders less susceptible than the latter disorders to the influences of changes in adult gender roles. That argument does not extend to generalized anxiety disorder, though, which has a similar age of onset distribution to MDD. [55] Why narrowing occurred for MDD but not GAD, then, is especially puzzling. Increased understanding of these specifications should be the subject of future theorizing and empirical investigation.

Acknowledgments

References

1. Kuehner C. Gender differences in unipolar depression: an update of epidemiological findings and possible explanations. Acta Psychiatr Scand 2003;108(3):163–174. [PubMed: 12890270]

2. Pigott TA. Gender differences in the epidemiology and treatment of anxiety disorders. J Clin Psychiatry 1999;60 (Suppl 18):4–15. [PubMed: 10487250]

3. Arnold LE. Sex differences in ADHD: conference summary. J Abnorm Child Psychol 1996;24(5): 555–569. [PubMed: 8956084]

4. Brady KT, Randall CL. Gender differences in substance use disorders. Psychiatr Clin North Am 1999;22(2):241–252. [PubMed: 10385931]

5. Keenan K, Loeber R, Green S. Conduct disorder in girls: a review of the literature. Clin Child Fam Psychol Rev 1999;2(1):3–19. [PubMed: 11324095]

6. Grigoriadis S, Robinson GE. Gender issues in depression. Ann Clin Psychiatry 2007;19(4):247–255. [PubMed: 18058282]

7. Lynch WJ, Roth ME, Carroll ME. Biological basis of sex differences in drug abuse: preclinical and clinical studies. Psychopharmacology (Berl) 2002;164(2):121–137. [PubMed: 12404074]

8. Nolen-Hoeksema S, Hilt L. Possible contributors to the gender differences in alcohol use and problems. J Gen Psychol 2006;133(4):357–374. [PubMed: 17128956]

9. Joyce PR, Oakley-Browne MA, Wells JE, Bushnell JA, Hornblow AR. Birth cohort trends in major depression: increasing rates and earlier onset in New Zealand. J Affect Disord 1990;18(2):83–89. [PubMed: 2137473]

10. Wickramaratne PJ, Weissman MM, Leaf PJ, Holford TR. Age, period and cohort effects on the risk of major depression: results from five United States communities. J Clin Epidemiol 1989;42(4):333– 343. [PubMed: 2723694]

11. Bloomfield K, Gmel G, Neve R, Mustonen H. Investigating Gender Convergence in Alcohol Consumption in Finland, Germany, The Netherlands, and Switzerland: A Repeated Survey Analysis. Subst Abus 2001;22(1):39–53. [PubMed: 12466668]

12. McPherson M, Casswell S, Pledger M. Gender convergence in alcohol consumption and related problems: issues and outcomes from comparisons of New Zealand survey data. Addiction 2004;99 (6):738–748. [PubMed: 15139872]

13. Wilsnack RW, Vogeltanz ND, Wilsnack SC, Harris TR, Ahlstrom S, Bondy S, Csemy L, Ferrence R, Ferris J, Fleming J, Graham K, Greenfield T, Guyon L, Haavio-Mannila E, Kellner F, Knibbe R, Kubicka L, Loukomskaia M, Mustonen H, Nadeau L, Narusk A, Neve R, Rahav G, Spak F, Teichman M, Trocki K, Webster I, Weiss S. Gender differences in alcohol consumption and adverse drinking consequences: cross-cultural patterns. Addiction 2000;95(2):251–265. [PubMed: 10723854]

14. Pape H, Hammer T, Vaglum P. Are "traditional" sex differences less conspicuous in young cannabis users than in other young people? J Psychoactive Drugs 1994;26(3):257–263. [PubMed: 7844655]

15. Thoits P. Multiple identities: examining gender and marital status differences in distress. Am Sociol Rev 1986;51:259–272.

16. Wauterickx N, Bracke P. Unipolar depression in the Belgian population: trends and sex differences in an eight-wave sample. Soc Psychiatry Psychiatr Epidemiol 2005;40(9):691–699. [PubMed: 16142512]

17. Kessler RC, Demler O, Frank RG, Olfson M, Pincus HA, Walters EE, Wang P, Wells KB, Zaslavsky AM. Prevalence and treatment of mental disorders, 1990 to 2003. N Engl J Med 2005;352(24):2515– 2523. [PubMed: 15958807]

18. Roos E, Lahelma E, Rahkonen O. Work-family conflicts and drinking behaviours among employed women and men. Drug Alcohol Depend 2006;83(1):49–56. [PubMed: 16300907]

19. Sachs-Ericsson N, Ciarlo JA. Gender, social roles, and mental health: an epidemiological perspective. Sex Roles 2000;43:605–628.

20. Weich S, Sloggett A, Lewis G. Social roles and gender difference in the prevalence of common mental disorders. Br J Psychiatry 1998;173:489–493. [PubMed: 9926077]

21. Kessler, RC.; Ustun, TB., editors. The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders. Cambridge, UK: Cambridge University Press; 2008.

22. 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 H, Yeh EK. Cross-national epidemiology of major depression and bipolar disorder. JAMA 1996;276 (4):293–299. [PubMed: 8656541]

23. World Bank. World Development Report 2004: Making Services Work for Poor People. Washington, DC: The International Bank for Reconstruction and Development/The World Bank; 2003.

24. Heeringa, SG.; Wells, JE.; Hubbard, F.; Mneimneh, Z.; Chiu, WT.; Sampson, NA.; Berglund, PA. Sample Designs and Sampling Procedures. In: Kessler, RC.; Ustun, TB., editors. The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders. Cambridge, UK: Cambridge University Press; 2008. p. 14-32.

25. Pennell, BE.; Mneimneh, Z.; Bowers, A.; Chardoul, S.; Wells, JE.; Viana, MC.; Dinkelmann, K.; Gebler, N.; Florescu, S.; He, Y.; Huang, Y.; Tomov, T.; Vilagut, G. Implementation of the World Mental Health Surveys. In: Kessler, RC.; Ustun, TB., editors. The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders. Cambridge, UK: Cambridge University Press; 2008. p. 33-57.

26. Kessler RC, Ustun TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). Int J Methods Psychiatr Res 2004;13(2):93–121. [PubMed: 15297906]

27. Harkness, J.; Pennell, BE.; Villar, A.; Gebler, N.; Aguilar-Gaxiola, S.; Bilgen, I. Translation procedures and translation assessment in the World Mental Health Survey Initiative. In: Kessler, RC.; Ustun, TB., editors. The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders. Cambridge, UK: Cambridge University Press; 2008. p. 91-113.

28. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, (DSM-IV), Fourth Edition. Washington, DC: American Psychiatric Association; 1994.

29. Haro JM, Arbabzadeh-Bouchez S, Brugha TS, de Girolamo G, Guyer ME, Jin R, Lepine JP, Mazzi F, Reneses B, Vilagut G, Sampson NA, Kessler RC. Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys. Int J Methods Psychiatr Res 2006;15(4):167–180. [PubMed: 17266013]

30. Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi E, Howes MJ, Jin R, Secnik K, Spencer T, Ustun TB, Walters EE. The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. Psychol Med 2005;35(2):245–256. [PubMed: 15841682]

31. Degenhardt L, Bohnert KM, Anthony JC. Case ascertainment of alcohol dependence in general population surveys: 'gated' versus 'ungated' approaches. Int J Methods Psychiatr Res 2007;16(3): 111–123. [PubMed: 17703472]

32. Degenhardt L, Bohnert KM, Anthony JC. Assessment of cocaine and other drug dependence in the general population: "gated" versus "ungated" approaches. Drug Alcohol Depend 2008;93(3):227–232. [PubMed: 18065161]

33. Degenhardt L, Cheng H, Anthony JC. Assessing cannabis dependence in community surveys: methodological issues. Int J Methods Psychiatr Res 2007;16(2):43–51. [PubMed: 17623384]

34. Knauper B, Cannell CF, Schwarz N, Bruce ML, Kessler RC. Improving the accuracy of major depression age of onset reports in the US National Comorbidity Survey. Int J Methods Psychiatr Res 1999;8(1):39–48.

35. Fischer AH, Rodriguez Mosquera PM, van Vianen AE, Manstead AS. Gender and culture differences in emotion. Emotion 2004;4(1):87–94. [PubMed: 15053728]

36. Schmitt DP. Sociosexuality from Argentina to Zimbabwe: A 48-nation study of sex, culture, and strategies of human mating. Behavioral and Brain Sciences 2005;28:247–275. [PubMed: 16201459]

37. Wood W, Eagly AH. A cross-cultural analysis of the behavior of women and men: implications for the origins of sex differences. Psychol Bull 2002;128(5):699–727. [PubMed: 12206191]

38. Lopez-Claros, A.; Zahidi, S. Women's Empowerment: Measuring the Global Gender Gap. Geneva: World Economic Forum; 2005.

39. Efron B. Logistic regression, survival analysis, and the Kaplan-Meier curve. J Am Stat Assoc 1988;83:414–425.

40. Raftery AE. Bayesian model selection in social research. Sociological Methodology 1995;25:111–163.

41. Wolter, KM. Introduction to Variance Estimation. New York, NY: Springer-Verlag; 1985.

42. SUDAAN: Professional Software for Survey Data Analysis. Research Triangle Park, NC: Research Triangle Institute; 2002. [computer program]. Version 8.0.1

43. Mplus. Los Angeles, CA: Muthen & Muthen; 2007. [computer program]. Version 5

44. Wilhelm K, Parker G, Geerligs L, Wedgwood L. Women and depression: a 30 year learning curve. Aust N Z J Psychiatry 2008;42(1):3–12. [PubMed: 18058438]

45. Murphy JM, Laird NM, Monson RR, Sobol AM, Leighton AH. Incidence of depression in the Stirling County Study: historical and comparative perspectives. Psychol Med 2000;30(3):505–514. [PubMed: 10883707]

46. Grant BF. Prevalence and correlates of drug use and DSM-IV drug dependence in the United States: results of the National Longitudinal Alcohol Epidemiologic Survey. J Subst Abuse 1996;8(2):195–210. [PubMed: 8880660]

47. Keyes KM, Grant BF, Hasin DS. Evidence for a closing gender gap in alcohol use, abuse, and dependence in the United States population. Drug Alcohol Depend 2008;93(1–2):21–29. [PubMed: 17980512]

48. Nelson CB, Heath AC, Kessler RC. Temporal progression of alcohol dependence symptoms in the U.S. household population: results from the National Comorbidity Survey. J Consult Clin Psychol 1998;66(3):474–483. [PubMed: 9642885]

49. Jenkins R. Sex differences in minor psychiatric morbidity. Psychol Med Monogr Suppl 1985;7:1–53. [PubMed: 3875115]

50. Celentano DD, McQueen DV. Alcohol consumption patterns among women in Baltimore. J Stud Alcohol 1984;45:355–358. [PubMed: 6332947]

51. Parker DA, Harford TC. Gender-role attitudes, job competition and alcohol consumption among women and men. Alcohol Clin Exp Res 1992;16(2):159–165. [PubMed: 1590535]

52. Rahav G, Wilsnack R, Bloomfield K, Gmel G, Kuntsche S. The influence of societal level factors on men's and women's alcohol consumption and alcohol problems. Alcohol Alcohol Suppl 2006;41 (1):i47–55. [PubMed: 17030503]

53. Wilsnack SC, Wilsnack RW. International gender and alcohol research: recent findings and future directions. Alcohol Res Health 2002;26(4):245–250. [PubMed: 12875033]

54. Ahlstrom S, Bloomfield K, Knibbe R. Gender Differences in Drinking Patterns in Nine European Countries: Descriptive Findings. Subst Abus 2001;22(1):69–85. [PubMed: 12466670]

55. Kessler, RC.; Aguilar-Gaxiola, S.; Alonso, J.; Angermeyer, MC.; Anthony, JC.; Berglund, PA.; Chatterji, S.; de Girolamo, G.; de Graaf, R.; Demyttenaere, K.; Gasquet, I.; Gluzman, S.; Gruber, MJ.; Gureje, O.; Haro, JM.; Heeringa, S.; Karam, A.; Kawakami, N.; Lee, S.; Levinson, D.; Medina-Mora, ME.; Oakley-Browne, MA.; Beth-Ellen Pennell, B-E.; Petukhova, M.; Posada-Villa, J.; Ruscio, A.; Stein, DJ.; Tsang, CHA.; Ustun, TB. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Mental Health Survey Initiative. In: Kessler, RC.; Ustun, TB., editors. The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders. Cambridge, UK: Cambridge University Press; 2008. p. 511-521.
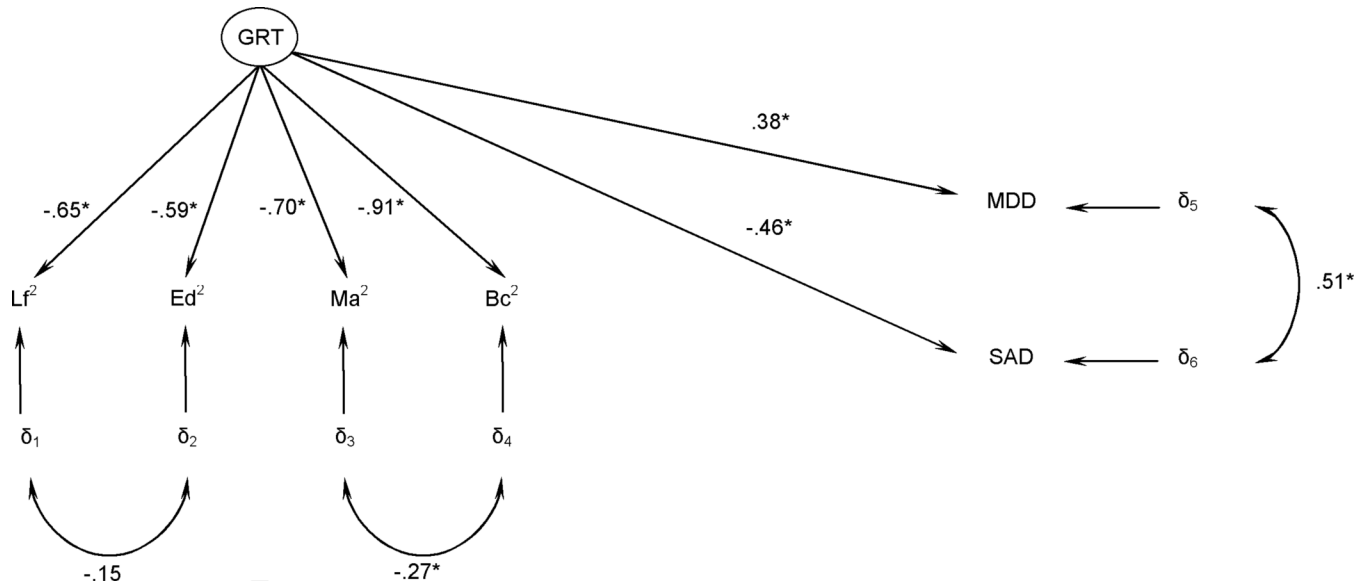
Figure 1.
Standardized[1] parameter estimates of the association between time-space variation in gender role traditionality (GRT) and the female-male odds-ratios of lifetime DSM-IV major depressive disorder (MDD) and substance abuse-dependence (SAD) in the 58 WMH cohort-by-country time-space subsamples

[1]Observed variables and GRT are all standardized to a mean of 0.0 and a standardized deviation of 1.0. See the text for a discussion of the transformations of model parameters into substantively meaningful matrices.

[2] Lf = Female to male percentage in labor force by age 35.

Ed= Female to male percentage reaching the median education level of the top quartile of earners.

Ma= Female to male median age at first marriage.

Bc= Percentage of women using birth control before the age of 25.

NIH-PA Author Manuscript

Table 1

WMH Sample Characteristics

Country	Survey[1]	Sample Characteristics[2]	Field Dates	Age Range	Sample Size			Response Rate
					Part I	Part II	Part II and Age ≤ 44[4]	
I. Developed countries								
Belgium 4.30	ESEMeD	Stratified multistage clustered probability sample of individuals residing in households from the national register of Belgium residents. Nationally representative (NR)	2001–2	18+	(2419)	(1043)	(486)	50.6
France	ESEMeD	Stratified multistage clustered sample of working telephone numbers merged with a reverse directory (for listed numbers). Initial recruitment was by telephone, with supplemental in-person recruitment in households with listed numbers. NR	2001–2	18+	(2894)	(1436)	(727)	45.9
Germany	ESEMeD	Stratified multistage clustered probability sample of individuals from community resident registries. NR	2002–3	18+	(3555)	(1323)	(621)	57.8
Israel	NHS	Stratified multistage clustered area probability sample of individuals from a national resident register	2002–4	21+	(4859)	--	--	72.6
Italy	ESEMeD	Stratified multistage clustered probability sample of individuals from municipality resident registries. NR	2001–2	18+	(4712)	(1779)	(853)	71.3
Japan	WMHJ2002-2003	Un-clustered two-stage probability sample of individuals residing in households in four metropolitan areas (Fukiage, Kushikino, Nagasaki, Okayama)	2002–3	20+	(2436)	(887)	(282)	56.4
Netherlands	ESEMeD	Stratified multistage clustered probability sample of individuals residing in households that are listed in municipal postal registries. NR	2002–3	18+	(2372)	(1094)	(516)	56.4

Country	Survey[1]	Sample Characteristics[2]	Field Dates	Age Range	Sample Size			Response Rate
					Part I	Part II	Part II and Age ≤ 44[4]	
New Zealand	NZMHS	Stratified multistage clustered area probability sample of household residents. NR	2004–5	18+	(12790)	(7312)	(4119)	73.3
Spain	ESEMeD	Stratified multistage clustered area probability sample of household residents. NR	2001–2	18+	(5473)	(2121)	(960)	78.6
United States	NCS-R	Stratified multistage clustered area probability sample of household residents. NR	2002–3	18+	(9282)	(5692)	(3197)	70.9
II. Developing countries								
Colombia	NSMH	Stratified multistage clustered area probability sample of household residents in all urban areas of the country (approximately 73% of the total national population)	2003	18–65	(4426)	(2381)	(1731)	87.7
Lebanon	LEBANON	Stratified multistage clustered area probability sample of household residents. NR	2002–3	18+	(2857)	(1031)	(595)	70.0
Mexico	M-NCS	Stratified multistage clustered area probability sample of household residents in all urban areas of the country (approximately 75% of the total national population).	2001–2	18–65	(5782)	(2362)	(1736)	76.6
South Africa	SASH	Stratified multistage clustered area probability sample of household residents. NR	2003–4	18+	(4351)	--	--	87.1
Ukraine	CMDPSD	Stratified multistage clustered area probability sample of household residents. NR	2002	18+	(4725)	(1720)	(541)	78.3

[1] ESEMeD (The European Study Of The Epidemiology Of Mental Disorders); NHS (Israel National Health Survey); WMHJ2002–2003 (World Mental Health Japan Survey); NZMHS (New Zealand Mental Health Survey); NCS-R (The US National Comorbidity Survey Replication).NSMH (The Colombian National Study of Mental Health); LEBANON (Lebanese Evaluation of the Burden of Ailments and Needs of the Nation); M-NCS (The Mexico National Comorbidity Survey); SASH (South Africa Health Survey); CMDPSD (Comorbid Mental Disorders during Periods of Social Disruption);

[2] Most WMH surveys are based on stratified multistage clustered area probability household samples in which samples of areas equivalent to counties or municipalities in the US were selected in the first stage followed by one or more subsequent stages of geographic sampling (e.g., towns within counties, blocks within towns, households within blocks) to arrive at a sample of households, in each of which a listing of household members was created and one or two people were selected from this listing to be interviewed. No substitution was allowed when the originally sampled household resident could not be interviewed. These household samples were selected from Census area data in all countries other than France (where telephone directories were used to select households) and the Netherlands (where postal registries were used to select households). Several WMH surveys (Belgium, Germany, Italy) used municipal resident registries to select respondents without listing households. The Japanese sample is the only totally un-

clustered sample, with households randomly selected in each of the four sample areas and one random respondent selected in each sample household. 12 of the 15 surveys are based on nationally representative (NR) household samples, while two others are based on nationally representative household samples in urbanized areas (Colombia, Mexico).

[3] The response rate is calculated as the ratio of the number of households in which an interview was completed to the number of households originally sampled, excluding from the denominator households known not to be eligible either because of being vacant at the time of initial contact or because the residents were unable to speak the designated languages of the survey. The weighted average response rate is 71.2%

[4] All countries with the exception of Ukraine (which was age restricted to ≤ 39) were age restricted to ≤ 44.

Table 2

Associations (odds-ratios) of gender with lifetime risk of DSM-IV mental disorders in the WMH surveys (n = 73,099)[1]

	All countries combined	Within-country OR variation				GxC interaction[2]	(n)[3]
		Range		% in dominant direction	% p ≤ .05 in dominant direction		
	F:M OR (95% CI)	Min	Max			χ^2	
I. Mood disorders							
Major depressive disorder	1.9* (1.8–2.0)	1.6	2.4	100.0	100.0	29.3*	(15)
Dysthymic disorder	1.9* (1.6–2.2)	1.3	3.8	100.0	70.0	13.0	(10)
Bipolar disorder	0.9 (0.8–1.0)	0.6	1.1	83.3	20.0	8.1	(6)
Any mood disorder	1.8* (1.7–1.8)	1.5	2.5	100.0	100.0	47.7*	(15)
II. Anxiety disorders							
Panic disorder	1.9* (1.7–2.2)	1.2	3.4	100.0	66.7	15.8	(12)
Generalized anxiety disorder	1.7* (1.5–1.9)	0.7	2.7	86.7	76.9	23.7*	(15)
Agoraphobia	2.0* (1.7–2.3)	1.4	4.6	100.0	62.5	18.5	(8)
Social phobia	1.3* (1.2–1.4)	1.1	2.0	100.0	46.2	13.8	(13)
Specific phobia	2.0* (1.9–2.2)	1.3	3.1	100.0	100.0	39.3*	(12)
Separation anxiety disorder	1.6* (1.4–1.8)	1.4	2.0	100.0	75.0	3.8	(4)
Post-traumatic stress disorder	2.6* (2.2–2.9)	1.3	6.4	100.0	78.6	30.6*	(14)
Any anxiety disorder	1.7* (1.6–1.8)	1.2	3.2	100.0	86.7	41.2*	(15)
III. Externalizing disorders							
Attention-deficit/hyperactivity disorder	0.6* (0.5–0.8)	0.3	0.6	100.0	20.0	6.9	(5)
Conduct disorder	0.5* (0.4–0.7)	0.3	0.6	100.0	100.0	29.5*	(3)
Intermittent explosive disorder	0.7* (0.6–0.8)	0.4	0.8	100.0	33.3	5.3	(6)
Oppositional-defiant disorder	0.8* (0.6–1.0)	0.5	0.8	100.0	33.3	18.2*	(3)
Any externalizing disorder	0.7* (0.6–0.8)	0.3	1.4	83.3	40.0	7.2	(12)
IV. Substance disorders							

| | All countries combined | Within-country OR variation | | | | | |
| | | Range | | % in dominant direction | % p ≤ .05 in dominant direction | GxC interaction[2] | |
	F:M OR (95% CI)	Min	Max			χ^2	(n)[3]
Alcohol abuse	0.2* (0.2–0.3)	0.1	0.4	100.0	100.0	128.6*	(15)
Alcohol dependence	0.3* (0.3–0.4)	0.1	0.4	100.0	100.0	95.6*	(11)
Drug abuse or dependence	0.4* (0.3–0.4)	0.1	0.4	100.0	100.0	35.3*	(5)
Any substance disorder	0.3* (0.2–0.3)	0.1	0.4	100.0	100.0	141.6*	(14)
V. Any disorder	1.1* (1.1–1.2)	0.7	2.2	66.7	70.0	111.5*	(15)

*
Significant at the .05 level, two-sided test

[1]
Based on discrete-time survival models that used respondent cohort (age at interview), sex, and person-year to predict first onset of each disorder both separately in each country and pooled across all countries.

[2]
GxC=gender-by-country; based on pooled cross-national survival model that included an interaction between gender and the dummy variables for country.

[3]
The number of countries differs across outcomes because not all disorders were assessed in all countries. The pooled results for any mood disorder, any anxiety disorder, any externalizing disorder, and any substance disorder pooled whatever disorders in the relevant category existed in the country across countries even though the set sometimes differed across countries.

Table 3

Interactions (odds-ratios) of gender with cohort in predicting lifetime risk of DSM-IV mental disorders in the WMH surveys (n = 73,099)[1]

	All countries combined	Within-country OR variation				
		Range				
	OR (95% CI)	Min	Max	% in dominant direction	% p ≤ .05 in dominant direction	(n)[2]
I. Mood disorders						
Major depressive disorder	0.9* (0.8–1.0)	0.6	1.1	73.3	36.4	(15)
Dysthymic disorder	0.9 (0.7–1.1)	0.2	1.5	50.0	25.0	(8)
Bipolar disorder	1.1 (0.9–1.3)	0.9	2.1	66.7	0.0	(6)
Any mood disorder	0.9* (0.8–0.9)	0.6	1.2	60.0	33.3	(15)
II. Anxiety disorders						
Panic disorder	1.0 (0.8–1.1)	0.3	3.5	58.3	0.0	(12)
Generalized anxiety disorder	1.1 (1.0–1.2)	0.4	1.7	60.0	11.1	(15)
Agoraphobia	1.0 (0.8–1.2)	0.7	1.6	50.0	0.0	(8)
Social phobia	1.0 (0.9–1.1)	0.7	2.3	58.3	14.3	(12)
Specific phobia	1.0 (0.9–1.1)	0.7	1.2	50.0	16.7	(12)
Separation anxiety disorder	0.9 (0.7–1.1)	0.7	1.3	50.0	0.0	(4)
Post-traumatic stress disorder	1.1 (1.0–1.3)	0.4	2.0	57.1	25.0	(14)
Any anxiety disorder	1.0 (0.9–1.0)	0.7	1.2	53.3	0.0	(15)
III. Externalizing disorders						
Attention-deficit/hyperactivity disorder	1.0 (0.7–1.5)	0.1	5.9	80.0	0.0	(5)
Conduct disorder	1.4 (0.8–2.2)	0.1	1.5	66.7	100.0	(3)
Intermittent explosive disorder	1.3* (1.1–1.5)	0.4	1.8	83.3	40.0	(6)
Oppositional-defiant disorder	0.9 (0.5–1.6)	0.7	0.9	100.0	0.0	(3)
Any externalizing disorder	1.1 (0.8–1.4)	0.2	6.5	58.3	28.6	(12)
IV. Substance disorders						
Alcohol abuse	1.5* (1.3–1.7)	0.7	4.0	85.7	66.7	(14)
Alcohol dependence	1.5* (1.2–1.8)	0.8	6.6	90.0	55.6	(10)

	All countries combined	Within-country OR variation				
		Range				
	OR (95% CI)	Min	Max	% in dominant direction	% p ≤ .05 in dominant direction	(n)[2]
Drug abuse or dependence	1.2 (0.9–1.5)	0.6	1.3	66.7	50.0	(3)
Any substance disorder	1.4 * (1.3–1.7)	0.7	5.0	85.7	66.7	(14)
V. Any disorder	1.0 (0.9–1.0)	0.8	1.7	60.0	0.0	(15)

*
Significant at the .05 level, two-sided test

[1]
Based on discrete-time survival models that used respondent cohort (age at interview), sex, person-year, and the interaction between sex and cohort to predict first onset of each disorder both separately in each country and pooled across all countries. Each respondent was classified as being in one of four cohorts defined based on age at interview (18–34, 35–49, 50–64, and 65+ years of age). A continuous 1–4 variable coded 1 for the earliest cohorts (i.e., ages 65+ at interview) and 4 for the most recent cohorts (i.e., 18–34 at interview) was used to create the interaction terms, where the 1–4 variable was multiplied by a 0–1 variable for gender (coded 1 for females and 0 for males). An interaction term with an OR significantly greater than 1.0 means that the F:M OR is significantly higher (i.e., higher relative prevalence among females than males) in more recent than older cohorts, while an OR less than 1.0 means that the F:M OR is significantly lower (ie., lower relative prevalence among females than males) nm more recent than older cohorts.

[2]
The number of countries differs across outcomes because not all disorders were assessed in all countries. The pooled results for any mood disorder, any anxiety disorder, any externalizing disorder, and any substance disorder pooled whatever disorders in the relevant category existed in the country across countries even though the set sometimes differed across countries.

Table 4

The distribution of the standardized[1] gender role traditionality composite measure across the WMH countries and cohorts

| | Cohorts | | | |
	18–34	35–49	50–64	65+
I. Developed countries				
Belgium	−3.7	−2.8	0.9	2.7
France	−3.5	−2.5	−0.7	0.9
Germany	−3.4	−3.4	0.9	1.6
Israel	−2.4	−0.4	−0.3	0.7
Italy	0.2	1.2	2.9	4.8
Japan	−4.4	0.1	−1.2	3.1
Netherlands	−5.1	−0.2	−1.9	1.7
New Zealand	−2.2	−0.9	−1.3	−1.3
Spain	−4.0	−0.7	3.3	3.4
U.S.	−3.5	−3.1	−2.3	2.0
II. Developing countries				
Colombia	−0.5	1.2	2.0	
Lebanon	0.2	4.4	8.3	8.3
Mexico	−0.6	2.0	3.1	.
South Africa	−0.8	1.1	1.2	2.2
Ukraine	−2.1	−4.8	0.2	−1.0

[1] Each indicator was standardized to have a mean of 0 and a variance of 1.0 across the sample of 58 time-space units. Each time-space unit was assigned equal weight in calculating the mean and variance. The composite was then created by summing the four standardized indicator scores and then standardizing the sum do have a mean of 0 zero and a variance of 1.0 across the sample of 58 time-space units

Table 5

Change in the associations of gender role traditionality with first onset of major depressive disorder and any substance disorder in models that either do or do not control for temporal or spatial variation

	Standardized regression coefficients linking gender role traditionality with			
	Major Depressive Disorder		**Any Substance Disorder**	
Controls	**Est**	**(se)**	**Est**	**(se)**
None[1]	.38[*]	(.19)	−.46[*]	(.19)
For temporal variation	.06	(.20)	−.44[*]	(.21)
For spatial variation	.60[*]	(.24)	−.20	(.18)

[*] Significant at the .05 level, two-sided test

[1] The coefficients in this row are those presented in Figure 1