Prevalence, Severity, and Unmet Need for Treatment of Mental Disorders in the World Health Organization World Mental Health Surveys

ALTHOUGH SURVEYS OF MENTAL disorders have been carried out since the end of World War II,1-3 cross-national comparisons were hampered by inconsistencies in diagnostic methods. This situation changed in the 1980s with the development of the Diagnostic Interview Schedule (DIS), the first psychiatric diagnostic interview designed for use by lay interviewers.4 The DIS was initially used in the US Epidemiologic Catchment Area (ECA) Study and subsequently in similar surveys carried out in other countries in the 1980s.5-8 The results were brought together in the early 1990s in a series of important cross-national articles that showed mental disorders to be highly prevalent.9-12 Indeed, prevalence of mental disorder was generally higher than that of any other class of chronic conditions.13,14 This was striking in light of research documenting that mental disorders have greater effects on role functioning than many serious chronic physical illnesses.13,15,16 A second generation of cross-national psychiatric surveys was carried out in the 1990s17-24 using a more elaborate interview, the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI).25 Although prevalence varied widely, from 4.3% in Shanghai to 26.4% in the United States, with an inter-quartile range (IQR) of 9.1%-16.9%. Between 33.1% (Colombia) and 80.9% (Nigeria) of 12-month cases were mild (IQR, 40.2%-53.3%). Serious disorders were associated with substantial role disability. Although disorder severity was correlated with probability of treatment in almost all countries, 35.5% to 50.3% of serious cases in developed countries and 76.3% to 85.4% in less-developed countries received no treatment in the 12 months before the interview. Due to the high prevalence of mild and subthreshold cases, the number of those who received treatment far exceeds the number of untreated serious cases in every country.

Conclusions Reallocation of treatment resources could substantially decrease the problem of unmet need for treatment of mental disorders among serious cases. Structural barriers exist to this reallocation. Careful consideration needs to be given to the value of treating some mild cases, especially those at risk for progressing to more serious disorders.

JAMA. 2004;291:2581-2590
www.jama.com

©2004 American Medical Association. All rights reserved.
countries could not afford to treat all the people with a mental disorder. Motivated by this concern, investigators performed secondary analyses of 2 US surveys, which concluded that up to half of 12-month mental disorders were mild. Another secondary analysis of CIDI surveys in 5 developed countries found a similar proportion of mild cases and showed that treatment was consistently correlated with severity. Between one third and two thirds of serious cases in these surveys nevertheless received no treatment.

The DIS and CIDI surveys had 3 limitations to analysis of severity and treatment. First, as they were designed to assess prevalence, not severity, the post hoc measures of severity used in secondary analyses of these surveys were weak. Second, the interviews did not include standardized treatment questions, thwarting valid cross-national comparisons of treatment. Third, the surveys were carried out mostly in developed countries, making it impossible to assess generalizability of results. WHO established the World Mental Health (WMH) Survey Consortium in 1998 to address such limitations. The CIDI was expanded to include detailed questions about disorder severities, impairments, and treatments.32 The CIDI was implemented in 28 countries around the world, including less-developed countries in each region of the world. The WMH surveys have now been completed in 14 countries, 6 of them less developed. This article is the first joint publication from these surveys. The focus is on aggregate estimates of 12-month prevalence, severity, and treatment.

METHODS

Samples

Fifteen surveys were carried out in 14 countries in the Americas (Colombia, Mexico, United States), Europe (Belgium, France, Germany, Italy, Netherlands, Spain, Ukraine), the Middle East and Africa (Lebanon, Nigeria), and Asia (Japan, separate surveys in Beijing and Shanghai in the People’s Republic of China). Six countries are classified by the World Bank as less developed (China, Colombia, Lebanon, Mexico, Nigeria, and Ukraine) and the others as developed. An effort was made to recruit as many countries as possible in the initiative. The final set was determined by availability of collaborators in the country who were able to obtain funding for the survey. All surveys were based on multistage household probability samples (Table 1). All interviews were carried out face-to-face by trained lay interviewers. The 6 Western European surveys were carried out jointly.35 Sample sizes range from 1663 (Japan) to 9282 (United States), with a total of 60463 participating adults. Response rates range from 45.9% (France) to 87.7% (Colombia), with a weighted average of 69.9%.

Internal subsampling was used to reduce respondent burden by dividing the interview into 2 parts. Part 1 included core diagnostic assessment. Part 2 included information about correlates and disorders of secondary interest. All respondents completed part 1. All part-1 respondents who met criteria for any disorder and a subsample of approximately 25% of others were administered part 2. The part-2 sample included 25828 respondents. Noncertainty part-2 respondents were weighted by the inverse of their probability of selection to adjust for differential sampling. Analyses in this article are based on this weighted part-2 sample. Additional weights were used to adjust for differential probabilities of selection within households and to match the samples to population sociodemographic distributions. The samples show substantial cross-national differences in age structure (younger in less-developed countries) and educational status (lower in less-developed countries). (Demographic distributions available on request.)

Training and Field Procedures

The central WMH staff trained bilingual supervisors in each country. Consistent interviewer training documents and procedures were used across surveys. The WHO translation protocol was used to translate instruments and training materials. Two surveys were carried out in bilingual form (Dutch and French in Belgium; Russian and Ukrainian in Ukraine). Others were carried out exclusively in the country’s official language (or, in Nigeria, in the Yoruba language that dominates in the region where the survey was carried out). Persons who could not speak these languages were excluded. Standardized descriptions of the goals and procedures of the study, data uses and protection, and the rights of respondents were provided in both written and verbal form to all predesignated respondents before obtaining verbal informed consent for participation in the survey. Quality control protocols described in more detail elsewhere were standardized across countries to check on interviewer accuracy and to specify data cleaning and coding procedures. The institutional review board of the organization that coordinated the survey in each country approved and monitored compliance with procedures for obtaining informed consent and protecting human subjects.

Measures

All surveys used the WMH-CIDI, a fully structured diagnostic interview, to assess disorders and treatment. Disorders considered herein include anxiety disorders (agoraphobia, generalized anxiety disorder, obsessive-compulsive disorder, panic disorder, posttraumatic stress disorder, social phobia, specific phobia), mood disorders (bipolar I and II disorders, dystymia, major depressive disorder), disorders that share a feature of problems with impulse control (bulimia, intermittent explosive disorder, and adult persistence of 3 childhood-adolescent disorders—attention-deficit/hyperactivity disorder, conduct disorder, and oppositional-defiant disorder—among respondents in the 18- to 44-year age range), and substance disorders (alcohol and drug abuse and dependence). Disorders were assessed...
Table 1. Sample Characteristics

<table>
<thead>
<tr>
<th>Country</th>
<th>Survey</th>
<th>Sample Characteristics*</th>
<th>Field Dates</th>
<th>Age Range, y</th>
<th>Sample Size</th>
<th>Response Rate, %†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>ESEMeD</td>
<td>Stratified multistage clustered probability sample of individuals residing in households from the national register of Belgium residents, nationally representative</td>
<td>2001-2002</td>
<td>≥18</td>
<td>2419</td>
<td>1043</td>
</tr>
<tr>
<td>Colombia</td>
<td>NSMHW</td>
<td>Stratified multistage clustered area probability sample of household residents in all urban areas of the country (approximately 73% of the total national population)</td>
<td>2003</td>
<td>18-65</td>
<td>4544</td>
<td>2442</td>
</tr>
<tr>
<td>France</td>
<td>ESEMeD</td>
<td>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, nationally representative</td>
<td>2001-2002</td>
<td>≥18</td>
<td>2894</td>
<td>1436</td>
</tr>
<tr>
<td>Germany</td>
<td>ESEMeD</td>
<td>Stratified multistage clustered probability sample of individuals from community resident registries, nationally representative</td>
<td>2002-2003</td>
<td>≥18</td>
<td>3555</td>
<td>1323</td>
</tr>
<tr>
<td>Italy</td>
<td>ESEMeD</td>
<td>Stratified multistage clustered probability sample of individuals from municipality resident registries, nationally representative</td>
<td>2001-2002</td>
<td>≥18</td>
<td>4712</td>
<td>1779</td>
</tr>
<tr>
<td>Japan</td>
<td>WMHJ2002-2003</td>
<td>Unclustered 2-stage probability sample of individuals residing in households in 4 metropolitan areas (Fukuage, Kushikino, Nagasaki, Oyayama)</td>
<td>2002-2003</td>
<td>≥20</td>
<td>1663</td>
<td>477</td>
</tr>
<tr>
<td>Lebanon</td>
<td>LNMHS</td>
<td>Stratified multistage clustered area probability sample of household residents, nationally representative</td>
<td>2002-2003</td>
<td>≥18</td>
<td>2856</td>
<td>1029</td>
</tr>
<tr>
<td>Mexico</td>
<td>M-NCS</td>
<td>Stratified multistage clustered area probability sample of household residents in all urban areas of the country (approximately 75% of the total national population)</td>
<td>2001-2002</td>
<td>18-65</td>
<td>5782</td>
<td>2362</td>
</tr>
<tr>
<td>Netherlands</td>
<td>ESEMeD</td>
<td>Stratified multistage clustered probability sample of individuals residing in households that are listed in municipal postal registries, nationally representative</td>
<td>2002-2003</td>
<td>≥18</td>
<td>2372</td>
<td>1094</td>
</tr>
<tr>
<td>Nigeria</td>
<td>NSMHW</td>
<td>Stratified multistage clustered area probability sample of household residents in the Yoruba-speaking Southwestern and North Central parts of the country (approximately 22% of the total national population)</td>
<td>2002</td>
<td>≥18</td>
<td>4985</td>
<td>1682</td>
</tr>
<tr>
<td>People’s Republic of China</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beijing</td>
<td>B-WMH</td>
<td>Stratified multistage clustered area probability sample of household residents in the Beijing metropolitan area</td>
<td>2002-2003</td>
<td>≥18</td>
<td>2633</td>
<td>914</td>
</tr>
<tr>
<td>Shanghai</td>
<td>S-WMH</td>
<td>Stratified multistage clustered area probability sample of household residents in the Shanghai metropolitan area</td>
<td>2002-2003</td>
<td>≥18</td>
<td>2568</td>
<td>714</td>
</tr>
<tr>
<td>Spain</td>
<td>ESEMeD</td>
<td>Stratified multistage clustered area probability sample of household residents, nationally representative</td>
<td>2001-2002</td>
<td>≥18</td>
<td>5473</td>
<td>2121</td>
</tr>
<tr>
<td>Ukraine</td>
<td>CMDPSD</td>
<td>Stratified multistage clustered area probability sample of household residents, nationally representative</td>
<td>2002</td>
<td>≥18</td>
<td>4725</td>
<td>1720</td>
</tr>
<tr>
<td>United States</td>
<td>NCS-R</td>
<td>Stratified multistage clustered area probability sample of household residents, nationally representative</td>
<td>2002-2003</td>
<td>≥18</td>
<td>9282</td>
<td>5692</td>
</tr>
</tbody>
</table>

Abbreviations: B-WMH, the Beijing World Mental Health Survey; CMDPSD, Comorbid Mental Disorders During Periods of Social Disruption; ESEMeD, the European Study of the Epidemiology of Mental Disorders; LNMHS, the Lebanese National Mental Health Survey; M-NCS, the Mexico National Comorbidity Survey; NCS-R, the US National Comorbidity Survey Replication; NSMHW, the Colombian National Study of Mental Health; NSMHW, the Nigerian Survey of Mental Health and Wellbeing; S-WMH, the Shanghai World Mental Health Survey; WMHJ2002-2003, World Mental Health Japan Survey.

*Most World Mental Health (WMH) surveys are based on stratified multistage clustered area probability household samples in which samples of areas equivalent to counties or municipalities in the United States were selected in the first stage followed by 1 or more subsequent stages of geographic sampling (eg, 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 1 or 2 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 (for which 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 unclustered sample, with households randomly selected in each of the 4 sample areas and 1 random respondent selected in each sample household. Nine of the 15 surveys are based on nationally representative household samples, while 2 others are based on nationally representative household samples in urbanized areas (Colombia, Mexico).

†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.

©2004 American Medical Association. All rights reserved.
using the definitions and criteria of the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)*. CIDI organic exclusion rules were imposed in making all diagnoses. Methodological evidence collected in the WHO-CIDI Field Trials and later clinical calibration studies showed that all the disorders considered were assessed with acceptable reliability and validity both in the original CIDI and in the original version of the WMH-CIDI. Studies of cross-national comparability in the validity of the WMH-CIDI are currently underway.

WMH-CIDI/DSM-IV disorders were classified as serious, moderate, or mild. Serious disorders were defined as one of the following: meeting criteria for bipolar I disorder or substance dependence with a physiological dependence syndrome; making a suicide attempt in conjunction with any other W MH-CIDI/DSM-IV disorder; reporting at least 2 areas of role functioning with severe role impairment due to a mental disorder in the disorder-specific Sheehan Disability Scales; or reporting overall functional impairment at a level consistent with a Global Assessment of Functioning of 50 or less in conjunction with any other W MH-CIDI/DSM-IV disorder. Respondents not classified as having a serious disorder were classified as moderate if interference was rated as at least moderate in any Sheehan Disability Scales domain or if the respondent had substance dependence without a physiological dependence syndrome. All other disorders were classified as mild. In an effort to validate severity ratings, respondents were asked how many days out of 365 in the past 12 months they were totally unable to carry out their normal daily activities because of each disorder. These reports were combined by assigning respondents who had more than 1 disorder to the highest number of days out of role reported for any single disorder.

Twelve-month treatment was assessed by asking respondents if they ever saw any of a long list of professionals either as an outpatient or inpatient for problems with emotions, nerves, mental health, or use of alcohol or drugs. Included were mental health professionals (eg, psychiatrist, psychologist), general medical professionals (eg, general practitioner, occupational therapist), religious counselors (eg, minister, sheikh), and

### Table 2. Twelve-Month Prevalence of World Mental Health Composite International Diagnostic Interview/Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition

<table>
<thead>
<tr>
<th>Country</th>
<th>Anxiety (%)</th>
<th>Mood (%)</th>
<th>Impulse-Control (%)</th>
<th>Substance (%)</th>
<th>Any (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Americas</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td>10.0 (8.4-11.7)</td>
<td>6.8 (6.0-7.7)</td>
<td>3.9 (3.2-4.7)</td>
<td>2.8 (2.0-3.7)</td>
<td>17.8 (16.1-19.5)</td>
</tr>
<tr>
<td>Mexico</td>
<td>6.8 (5.6-7.9) †</td>
<td>4.8 (4.0-5.6)</td>
<td>1.3 (0.9-1.8)</td>
<td>2.5 (1.8-3.3)</td>
<td>12.2 (10.5-13.80)</td>
</tr>
<tr>
<td>United States</td>
<td>18.2 (16.9-19.5)</td>
<td>9.6 (6.8-10.4)</td>
<td>6.8 (5.9-7.8)</td>
<td>3.8 (3.2-4.5)</td>
<td>26.4 (24.7-28.0)</td>
</tr>
<tr>
<td><strong>Europe</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>6.9 (4.5-9.4)</td>
<td>6.2 (4.8-7.6)§</td>
<td>1.0 (0.3-1.8)</td>
<td>1.2 (0.6-1.9) ††</td>
<td>12.0 (9.6-14.3)</td>
</tr>
<tr>
<td>France</td>
<td>12.0 (9.8-14.2)</td>
<td>8.5 (6.4-10.6)§</td>
<td>1.4 (0.7-2.0)</td>
<td>0.7 (0.3-1.2) ††</td>
<td>18.4 (15.3-21.5)</td>
</tr>
<tr>
<td>Germany</td>
<td>6.2 (4.7-7.8)</td>
<td>3.6 (2.8-4.3)§</td>
<td>0.3 (0.1-0.5)</td>
<td>1.1 (0.4-1.7) ††</td>
<td>9.1 (7.3-10.8)</td>
</tr>
<tr>
<td>Italy</td>
<td>5.8 (4.5-7.1)</td>
<td>3.8 (3.1-4.5)§</td>
<td>0.3 (0.1-0.5)</td>
<td>0.1 (0.0-0.2)‡‡</td>
<td>8.2 (6.7-9.7)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>8.8 (6.6-11.0)</td>
<td>6.9 (4.1-7.9)‡‡</td>
<td>1.3 (0.4-2.2)</td>
<td>3.0 (0.7-5.2) ††</td>
<td>14.9 (12.2-17.6)</td>
</tr>
<tr>
<td>Spain</td>
<td>5.9 (4.5-7.3)</td>
<td>4.9 (4.0-5.8)§</td>
<td>0.5 (0.2-0.8)</td>
<td>0.3 (0.0-0.5) ††</td>
<td>9.2 (7.8-10.6)</td>
</tr>
<tr>
<td>Ukraine</td>
<td>7.1 (5.6-8.6) ††</td>
<td>9.1 (7.3-10.9)§</td>
<td>3.2 (2.4-4.0)‡‡</td>
<td>6.4 (4.8-8.1) ††</td>
<td>20.5 (17.7-23.2)</td>
</tr>
<tr>
<td><strong>Middle East and Africa</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lebanon</td>
<td>11.2 (8.9-13.5)</td>
<td>6.6 (4.9-8.2)</td>
<td>1.7 (0.8-2.6)‡‡</td>
<td>1.3 (0.0-2.8)</td>
<td>16.9 (13.6-20.2)</td>
</tr>
<tr>
<td>Nigeria</td>
<td>3.3 (2.4-4.2)</td>
<td>0.8 (0.5-1.0)</td>
<td>0.0 (0.0-0.1)‡‡</td>
<td>0.8 (0.3-1.2)</td>
<td>4.7 (3.6-5.8)</td>
</tr>
<tr>
<td><strong>Asia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>5.3 (3.5-7.0)</td>
<td>3.1 (2.2-4.1)</td>
<td>1.0 (0.4-1.5) #<em>#</em>††</td>
<td>1.7 (0.3-3.0)</td>
<td>8.8 (6.4-11.2)</td>
</tr>
<tr>
<td>People’s Republic of China</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beijing</td>
<td>3.2 (1.8-4.6)†</td>
<td>2.5 (1.5-3.4)</td>
<td>2.6 (1.3-3.9)‡§#**</td>
<td>2.6 (1.2-3.9)</td>
<td>9.1 (6.0-12.1)</td>
</tr>
<tr>
<td>Shanghai</td>
<td>2.4 (0.9-3.9)</td>
<td>1.7 (0.6-2.9)</td>
<td>0.7 (0.4-1.1)‡§#**</td>
<td>0.5 (0.3-0.6)</td>
<td>4.3 (2.7-5.9)</td>
</tr>
</tbody>
</table>

*Anxiety disorders include agoraphobia, generalized anxiety disorder, obsessive-compulsive disorder, panic disorder, posttraumatic stress disorder, social phobia, and specific phobia. Mood disorders include bipolar I and II disorders, dysthymia, and major depressive disorder. Impulse-control disorders include bulimia, intermittent explosive disorder, and reported persistence in the past 12 months of symptoms of 5 child-adolescent disorders (attention-deficit hyperactivity disorder, conduct disorder, and oppositional-defiant disorder). Substance disorders include alcohol or drug abuse or dependence. In the case of substance dependence, respondents who met full criteria at some time in their life and who continue to have any symptoms are considered to have 12-month dependence even if they currently do not meet full criteria for the disorder. Organic exclusions were made as specified in the Diagnostic and Statistical Manual of Mental Health Disorders, Fourth Edition, but diagnostic hierarchy rules were not used.

†Obsessive-compulsive disorder was not assessed.

‡Specific phobia was not assessed.

§Bipolar disorders were not assessed.

¶Intermittent explosive disorder was not assessed.

#Bulimia was not assessed.

§Bipolar disorders were not assessed.

**Oppositional-defiant disorder was not assessed.

††Conduct disorder was not assessed.

¶¶Only alcohol abuse and dependence were assessed. No assessment was made of other drug abuse or dependence.
traditional healers (eg, herbalist, spiritualist). The list varied across countries depending on local circumstances. We focus herein on 12-month treatment by either a mental health professional or general medical professional.

**Analysis Methods**

Data are reported on prevalence, severity, and associations of severity with days out of role and with treatment. Simple cross-tabulations were used to calculate prevalence and severity. Associations of severity with days out of role and with treatment were examined using analyses of variance. Confidence intervals were estimated using the Taylor Series method with SUDAAN software to adjust for clustering and weighting. Multivariate tests were made using Wald χ² and F tests computed from design-adjusted coefficient variance–covariance matrices. Statistical significance was based on 2-sided tests evaluated at the .05 level of significance.

## Results

### Prevalence

Overall prevalence varies widely (Table 2), from 4.3% in Shanghai to 26.4% in the United States, with a 9.1% to 16.9% inter-quartile range (IQR, the range after excluding the highest and lowest 4 surveys). Anxiety disorders are the most common disorders in all but 1 country (higher prevalence of mood disorders in Ukraine), with prevalence in the range 2.4% to 18.2% (IQR, 5.8%-8.8%). Mood disorders are next most common in all but 2 countries (equal or higher prevalence of substance disorders in Nigeria and Beijing), with prevalence in the range 0.8% to 9.6% (IQR, 3.6%-6.8%). Substance disorders (12-month prevalence, 0.1%-6.4%; IQR, 0.8%-2.6%) and impulse-control disorders (12-month prevalence, 0.0%-6.8%; IQR, 0.7%-1.7%) are consistently less prevalent across the surveys. If we use the terms high and low to refer to the 5 highest and 5 lowest prevalence estimates in each column of the table, the United States and Colombia have consistently high prevalence estimates across all classes of disorder, the Netherlands and Ukraine are high on 3 of 4, Nigeria and Shanghai are consistently low, and Italy is low on 3 of 4.

### Severity

The proportions of the samples (Table 3) with either a serious disorder (0.4%-7.7%; IQR, 1.1%-3.7%) or a moderate disorder (0.5%-9.4%; IQR, 2.9%-6.1%) are generally smaller than the proportions with a mild disorder (1.8%-9.7%; IQR, 4.5%-6.4%). The proportion of disorders classified as mild is substantial: from 33.1% in Colombia to 80.9% in Nigeria (IQR, 40.2-%53.3%). The severity distribution among cases varies significantly across countries (χ² = 193.9, P < .001), with severity not strongly related either to region or to development status. There are substantial positive associations, however, between overall prevalence of any disorder and both the proportion of cases classified as serious (Pearson r = 0.56; P = .03) and the proportion of cases classified as either serious or moderate (Pearson r = 0.51; P = .05).

### Severity and Impairment

The severity classification was validated by documenting a statistically significant monotonic association between severity and days out of role in all but 2 surveys (Table 4). Respondents with serious disorders in most surveys reported at least 30 days in the past year when they were unable to carry out usual activities because of these disorders (IQR, 31.3-33.7 days; IQR, 9.2-18.8 days). Even the means for moderate disorders are larger than those found in previous research to be associated with most serious chronic physical disorders.

### Severity and Treatment

The proportion of respondents who received health care treatment for emotional or substance-use problems during the 12 months before the WMH interview varies widely across surveys (Table 5), from a low of 0.8% in Nigeria to a high of 94.3% in France.
### Table 4. Association Between Severity of 12-Month World Mental Health–Composite International Diagnostic Interview/Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition Disorders and Days Out of Role

<table>
<thead>
<tr>
<th>Country</th>
<th>Serious Mean (95% CI)</th>
<th>Moderate Mean (95% CI)</th>
<th>Mild Mean (95% CI)</th>
<th>F&lt;sub&gt;3, n-3&lt;/sub&gt; Wald F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Americas</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td>28.0 (13.8-42.3)</td>
<td>6.0 (2.9-9.1)</td>
<td>0.2 (0.1-0.3)</td>
<td>17.5*</td>
</tr>
<tr>
<td>Mexico</td>
<td>26.6 (17.0-36.1)</td>
<td>7.4 (3.9-10.8)</td>
<td>1.8 (0.0-4.5)</td>
<td>13.4*</td>
</tr>
<tr>
<td>United States</td>
<td>66.9 (56.0-77.8)</td>
<td>10.6 (7.7-13.5)</td>
<td>0.7 (0.1-1.3)</td>
<td>74.9*</td>
</tr>
<tr>
<td>Europe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>32.9 (7.8-58.0)</td>
<td>26.4 (5.3-47.5)</td>
<td>2.8 (0.0-8.1)</td>
<td>3.1</td>
</tr>
<tr>
<td>France</td>
<td>94.7 (46.5-142.9)</td>
<td>9.2 (4.6-13.7)</td>
<td>1.0 (0.0-3.0)</td>
<td>15.7*</td>
</tr>
<tr>
<td>Germany</td>
<td>84.6 (35.7-133.5)</td>
<td>13.4 (4.3-22.4)</td>
<td>0.3 (0.0-0.7)</td>
<td>10.1*</td>
</tr>
<tr>
<td>Italy</td>
<td>206.4 (114.9-297.9)</td>
<td>33.7 (13.0-54.4)</td>
<td>3.6 (1.8-5.5)</td>
<td>15.0*</td>
</tr>
<tr>
<td>Netherlands</td>
<td>123.2 (73.7-172.7)</td>
<td>13.4 (0.0-27.5)</td>
<td>1.3 (0.0-3.2)</td>
<td>13.9*</td>
</tr>
<tr>
<td>Spain</td>
<td>81.4 (33.5-129.2)</td>
<td>10.5 (4.8-16.1)</td>
<td>0.1 (0.0-0.3)</td>
<td>17.1*</td>
</tr>
<tr>
<td>Ukraine</td>
<td>38.1 (23.1-53.0)</td>
<td>18.8 (14.3-23.4)</td>
<td>0.7 (0.0-1.9)</td>
<td>42.0*</td>
</tr>
<tr>
<td>Middle East and Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lebanon</td>
<td>37.1 (19.7-54.6)</td>
<td>17.9 (7.5-28.3)</td>
<td>0.8 (0.0-1.9)</td>
<td>13.9*</td>
</tr>
<tr>
<td>Nigeria</td>
<td>15.2 (0.8-29.6)</td>
<td>18.8 (0.0-40.3)</td>
<td>0.6 (0.0-1.6)</td>
<td>3.0</td>
</tr>
<tr>
<td>Asia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>32.1 (0.0-65.6)</td>
<td>6.3 (1.6-11.0)</td>
<td>0.1 (0.0-0.2)</td>
<td>6.8*</td>
</tr>
<tr>
<td>People’s Republic of China</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beijing</td>
<td>25.9 (7.4-44.3)</td>
<td>23.1 (4.7-41.6)</td>
<td>0.4 (0.0-0.8)</td>
<td>7.2*</td>
</tr>
<tr>
<td>Shanghai</td>
<td>47.1 (13.6-80.7)</td>
<td>4.1 (0.0-10.3)</td>
<td>1.0 (0.0-2.5)</td>
<td>4.5*</td>
</tr>
</tbody>
</table>

*Significant association between severity and days out of role at the .05 level.

### Table 5. Association of 12-Month World Mental Health–Composite International Diagnostic Interview/Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition Disorder Severity With Health Care Treatment

<table>
<thead>
<tr>
<th>Country</th>
<th>% Serious (95% CI)</th>
<th>Moderate (95% CI)</th>
<th>Mild (95% CI)</th>
<th>None (95% CI)</th>
<th>Total (95% CI)</th>
<th>χ² Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Americas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td>23.7 (15.2-32.3)</td>
<td>11.5 (6.6-16.5)</td>
<td>8.4 (4.5-12.4)</td>
<td>3.0 (1.9-4.0)</td>
<td>5.0 (3.8-6.1)</td>
<td>41.1†</td>
</tr>
<tr>
<td>Mexico</td>
<td>20.2 (12.7-27.8)</td>
<td>18.6 (12.5-24.8)</td>
<td>10.2 (5.5-14.9)</td>
<td>2.6 (1.9-3.4)</td>
<td>4.2 (3.3-5.1)</td>
<td>73.4†</td>
</tr>
<tr>
<td>United States</td>
<td>52.3 (48.5-56.1)</td>
<td>34.1 (30.9-37.4)</td>
<td>22.5 (19.0-26.1)</td>
<td>8.1 (7.1-9.2)</td>
<td>15.3 (14.1-16.5)</td>
<td>278.4†</td>
</tr>
<tr>
<td>Europe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>53.9 (25.2-82.5)</td>
<td>50.0 (35.8-64.2)</td>
<td>28.2 (14.9-41.4)</td>
<td>7.2 (4.2-10.1)</td>
<td>11.0 (7.6-14.4)</td>
<td>68.0†</td>
</tr>
<tr>
<td>France</td>
<td>63.3 (38.6-88.1)</td>
<td>35.7 (21.4-49.9)</td>
<td>22.3 (15.8-28.9)</td>
<td>7.8 (5.7-10.0)</td>
<td>12.4 (10.2-14.6)</td>
<td>29.7†</td>
</tr>
<tr>
<td>Germany</td>
<td>49.7 (26.6-72.8)</td>
<td>30.5 (18.5-42.5)</td>
<td>27.9 (14.5-41.3)</td>
<td>5.4 (3.5-7.2)</td>
<td>7.8 (6.0-9.5)</td>
<td>37.9†</td>
</tr>
<tr>
<td>Italy</td>
<td>30.5 (19.3-41.7)</td>
<td>18.9 (11.3-26.6)</td>
<td>2.4 (1.6-3.2)</td>
<td>4.5 (3.6-5.5)</td>
<td>64.2†</td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>50.2 (29.5-70.8)</td>
<td>35.0 (15.7-54.2)</td>
<td>26.5 (15.6-37.4)</td>
<td>6.9 (4.4-9.4)</td>
<td>10.7 (8.1-13.2)</td>
<td>46.6†</td>
</tr>
<tr>
<td>Spain</td>
<td>64.5 (49.2-79.7)</td>
<td>37.9 (26.8-49.0)</td>
<td>35.2 (23.8-46.6)</td>
<td>4.0 (3.1-5.0)</td>
<td>7.3 (6.2-8.4)</td>
<td>152.1†</td>
</tr>
<tr>
<td>Ukraine</td>
<td>19.7 (13.9-25.6)</td>
<td>17.1 (9.7-24.4)</td>
<td>7.1 (1.2-13.0)</td>
<td>2.6 (1.5-3.8)</td>
<td>4.9 (3.5-6.3)</td>
<td>42.8†</td>
</tr>
<tr>
<td>Middle East and Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lebanon</td>
<td>14.6 (5.8-23.4)</td>
<td>9.7 (2.6-16.7)</td>
<td>4.5 (0.6-8.5)</td>
<td>2.6 (1.3-3.9)</td>
<td>3.7 (2.4-5.0)</td>
<td>14.6†</td>
</tr>
<tr>
<td>Nigeria</td>
<td>. . .</td>
<td>10.3 (3.7-17.0)</td>
<td>0.3 (0.0-0.5)</td>
<td>0.8 (0.4-1.1)</td>
<td>11.3†</td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>16.7 (4.5-28.9)</td>
<td>11.2 (0.1-22.3)</td>
<td>4.7 (2.6-6.9)</td>
<td>5.7 (3.7-7.7)</td>
<td>21.3†</td>
<td></td>
</tr>
<tr>
<td>People’s Republic of China</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beijing</td>
<td>. . .</td>
<td>11.9 (0.0-26.2)</td>
<td>2.0 (0.0-4.8)</td>
<td>2.4 (1.0-3.8)</td>
<td>2.7 (1.3-4.2)</td>
<td>3.1</td>
</tr>
<tr>
<td>Shanghai</td>
<td>. . .</td>
<td>0.5 (0.0-1.7)</td>
<td>2.3 (0.6-4.1)</td>
<td>3.1 (1.1-5.0)</td>
<td>6.2</td>
<td></td>
</tr>
</tbody>
</table>

*Ellipses indicate that the results were not reported because of sparse data (<30 respondents at the severity level in the survey).†Significant association between severity and probability of treatment at the .05 level.
majory of people in treatment in each country are either noncases or mild cases. (Results available on request.) These will be referred to for the remainder of this article as subthreshold cases. We also examined the associations of severity with 2 indicators of treatment intensity among people in health care treatment: being seen in the specialty mental health sector rather than exclusively in the general medical sector and number of visits in the 12 months before the interview. Statistical power was low in these analyses because of the small numbers of treated cases with serious conditions in most countries. Nevertheless, there was a clear trend in the vast majority of countries for severity to be positively related both to proportional treatment in the specialty sector and to number of visits, with the highest scores on each consistently found among serious cases. (Results available on request.)

Even though a dose-response relationship exists between severity and probability of treatment in virtually all countries, substantial proportions of serious cases receive no treatment. This is true even in developed countries, where 35.5% to 50.3% of serious cases were untreated in the health care sector in the year before the interview. The situation is even worse in less-developed countries, where 76.3% to 85.4% of serious cases received no treatment. This is especially striking in light of the fact that such a high proportion of treatment in all countries is devoted to subthreshold cases. It is interesting to note that the 3 surveys with the highest overall 12-month prevalence estimates (United States, Ukraine, and Colombia) also had 3 of the 4 lowest proportions of treatment devoted to subthreshold cases (52%-59%). In comparison, the 3 Asian surveys, all of which had quite low overall 12-month prevalence estimates, had the 3 highest proportions of treatment devoted to subthreshold cases (71%-85%).

**COMMENT**

An important limitation of the WMH surveys is their wide variation in response rate. In addition, some of the surveys had response rates below normally accepted standards. We attempted to adjust for differential response to the extent possible by poststratification, but this only deals with a limited type of bias. If response is related to mental illness, severity, or treatment in ways that cannot be corrected by simple sociodemographic adjustment, cross-national comparisons will be distorted.

A related limitation is that the Western European surveys, which were fielded before any of the other WMH surveys, experienced a number of difficulties in survey implementation, largely skip logic errors, that subsequent surveys avoided because they were resolved while carrying out the Western European surveys. As a result, these early surveys had much more item-missing data than later surveys, which led to underestimation of severity of some disorders because the Sheehan Disability Scales were sometimes mistakenly skipped.

An added complication was that various of the WMH surveys deleted disorders that were thought to have low relevance in their countries, leading to inconsistency in completeness of coverage. We investigated the implications of this variation by replicating analyses using only the disorders that were assessed in all surveys. Although basic patterns of association remained stable in these revised analyses (results available on request), it is still possible that some findings were sensitive to differential exclusion of some disorders in particular countries.

Another limitation is that schizophrenia and other nonaffective psychoses, although important mental disorders, were not included in the core WMH assessment because previous validation studies showed they are dramatically overestimated in lay-administered interviews like the WMH-CIDI.44-49 These same studies also showed, however, that the vast majority of respondents with clinician-diagnosed nonaffective psychoses meet criteria for CIDI anxiety, mood, or substance disorders and are consequently captured as cases even if nonaffective psychoses are not assessed.

A final noteworthy limitation is that the WMH-CIDI might vary in accuracy across countries. Although the previous methodological studies that were cited in the measurement section documented that earlier versions of the CIDI had acceptable concordance with blind clinical reinterviews, these studies were carried out largely in developed Western countries. Performance of the WMH-CIDI could be worse in other parts of the world either because the concepts and phrases used to describe mental syndromes are less consonant with cultural concepts than in developed Western countries or because absence of a tradition of free speech and anonymous public opinion surveying causes greater reluctance to admit emotional or substance-abuse problems than in developed Western countries.

Clinical reappraisal studies are currently underway in both developed and less developed WMH countries in all major regions of the world to evaluate the issue of cross-national differences in WMH-CIDI diagnostic validity. Even before completing these studies, though, some patterns in the data (eg, the much lower estimated rate of alcoholism in Ukraine than expected from administrative data documenting an important role of alcoholism in mortality in that country)50 raise concerns about differential validity. The most striking such pattern is that countries with the lowest disorder prevalence estimates have the highest proportion of respondents in treatment who are subthreshold cases. This pattern could very well reflect greater underestimation of disorders in countries with the lowest prevalence estimates.

Within the context of these limitations, the WMH results are consistent with those of earlier surveys in showing that mental disorders are highly prevalent,8-12 often are associated with serious role impairment,15,16,51 and often go untreated.27,28,52 We also found substantial cross-national variation in these results. Two broad patterns consistent with previous research are that...
prevalence is low in Asian countries and that treatment is low in less developed countries. There are so many idiosyncratic substantive and methodological factors that might contribute to these and other cross-national differences that it is more profitable to focus on consistency rather than on differences, at least in this initial report of broad WMH findings. It is noteworthy in this regard that prevalence and severity estimates are likely to be conservative, for previous methodological studies have shown that survey respondents tend to have significantly higher rates and severity of mental illness than respondents. The estimates of proportional treatment, in comparison, are likely to be downwardly biased because hospitalized patients were excluded from the surveys.

We found that disorder severity is strongly related to treatment in all countries. This finding is consistent with 2 previous large-scale survey investigations of the relationship between severity and treatment. Correction for response bias would likely strengthen this relationship. The most reasonable interpretation is that demand for treatment is related to severity, presumably mediated by distress and impairment. A question could be raised whether this is merely a matter of demand or whether the treatment system is also more receptive to more severe cases. Some indirect indication of system responsiveness can be gleaned from the findings (available on request) that treatment intensity, as indicated by proportional treatment in the specialty sector and number of visits, is greater for serious than for other treated cases in most WMH countries.

Despite this evidence of rationality in treatment resource allocation, we found that 35.5% to 50.3% of serious cases in treatment more than the proportion of subthreshold cases in treatment, leading to this pattern becoming even stronger. The fact that many people with subthreshold disorders are treated while many with serious disorders are not shows that unmet need for treatment among serious cases is not merely a matter of limited treatment resources but that misallocation of treatment resources is also involved.

A major practical difficulty in rationalizing allocation of treatment resources is that system barriers constrain reallocation options. This is especially true in a decentralized system like in the United States. For example, there is no obvious mechanism by which constraining access to psychotherapy among middle-class persons with mild mental disorders in the United States would result in an increase in treatment of low-income people with serious mental illness. Another complexity is that misallocation of treatment resources is partly due to differences in perceived need for treatment that are unrelated to objective severity and to differences in access associated with insurance coverage and financial resources.

Report comparing the mental health care delivery systems in the United States and Ontario showed that these 2 systems differ along exactly these lines. A higher proportion of people with serious mental illness were treated in Ontario than were treated in the United States because of lower constraint on access among persons unable to pay in Ontario than were able to pay in the United States while a higher proportion of mild cases were treated in the United States than Ontario because of significantly higher perceived need for treatment among insured middle-class people with mild disorders in the United States than in Ontario. Although a number of structural possibilities exist to modify constraints on access, it is unclear how perceived need could be modified to align demand with true need for treatment.

A final complexity in reallocating treatment resources is that optimal allocation rules are not obvious. The simplistic strategy of not treating any mild disorders is almost certainly suboptimal because we know that many people with mild disorders, especially young people, go on to develop serious mental disorders. To the extent that early intervention can prevent progression, early treatment of mild cases might be cost effective. It is difficult to act on this insight, however, because we lack good information either about the characteristics of mild cases that predict risk of progression to more serious disorders or about the effectiveness of interventions for mild cases in preventing this progression. A new focus on the development and evaluation of secondary prevention programs for the early treatment of mild cases is needed to guide rationalization of treatment resource allocation.


Author Contributions: Dr Kessler, as principal investigator, had full access to all of the data in this study and takes responsibility for the integrity of the data and the accuracy of the data analysis.


©2004 American Medical Association. All rights reserved.
PREVALENCE AND SEVERITY OF MENTAL DISORDERS

51. Kessler RC, Frank RG. The impact of psychiatric disorders on work loss days. Psychol Med. 1997;27:861-873.