A 40-Year Perspective on the Prevalence of Depression

The Stirling County Study

Jane M. Murphy, PhD; Nan M. Laird, PhD; Richard R. Monson, MD, DSc; Arthur M. Sobol, MA; Alexander H. Leighton, MD

Background: According to epidemiologic studies that use recall of lifetime episodes, the prevalence of depression is increasing. This report from the Stirling County Study compares rates of current depression among representative samples of adults from a population in Atlantic Canada.

Methods: Sample sizes were 1003, 1201, and 1396 in 1952, 1970, and 1992, respectively. The depression component of the study’s method, the DPAX (DP for depression and AX for anxiety), was employed. The original procedure (DPAX-1) was applied in all years. A revision (DPAX-2) was used in 1970 and 1992. The Diagnostic Interview Schedule (DIS) was also used in 1992.

Results: With the DPAX-1, the overall prevalence of current depression was steady at 5% over the 2 early samples but declined in 1992 because of vernacular changes referring to dysphoria. The DPAX-2 gave a stable overall prevalence of 5% in the 2 recent samples, but indicated that women and younger people were at greater risk in 1992 than in 1970. The DIS, like the DPAX-2, found a current 1992 rate of 5% for major depressive episodes combined with dysthymia. Recalled lifetime rates using the DIS showed the same profile interpreted in other studies as suggesting an increase in depression over time.

Conclusions: Three samples over a 40-year period showed a stable current prevalence of depression using the DPAX methods that was comparable in 1992 with the current rates using the DIS. This casts doubt on the interpretation that depression is generally increasing. Within the overall steady rate observed in this study, historical change was a matter of redistribution by sex and age, with a higher rate among younger women being of recent origin.

Arch Gen Psychiatry. 2000;57:209-215

See also pages 217, 223, 227, and 230

This report takes a different approach, in that we focus on what subjects said about their current moods, symptoms, and impairment when interviewed as members of representative samples from the same general population at 3 points (1952, 1970, and 1992), thus providing a population perspective that makes long-term recall unnecessary. As in an earlier report that dealt with 1952 and 1970, we use the study’s traditional methods for diagnosing disorders.¹⁴⁻¹⁶ We added the Diagnostic Interview Schedule (DIS) of the ECA for the 1992 survey so that comparisons could be drawn with other studies.¹⁷

The research site, given the pseudonym Stirling County and located in Canada on the eastern side of the Gulf of Maine, was selected for its socioeconomic and cultural diversity.¹⁸⁻²⁰ The population numbered 19,989 in the early 1950s and 21,255 in the 1990s. As a region, it has experienced the social trends that characterize North America generally. These include the social advances of a rising standard of living, improved health care delivery, and increased educational opportunities. The changing times have also brought disadvantages, such as racial tension and increases in crime, including illegal drug use; loosening of family ties; and weakening of religious values. The economic picture has fluctuated, rising to a
PARTICIPANTS AND METHODS
SAMPLING

The sampling procedures for 1932 and 1970 have been described previously. \[14,19,21\] The following points are relevant to the sample of 1992: (1) Research maps and the census were updated and continued to provide the means for selecting and locating subjects. (2) The 1952 sample was a selection of heads of households, while the samples of 1970 and 1992 were selections of all adults 18 years and older, among whom heads of households were designated. (3) Communities of special interest because of socioeconomic and cultural contrasts were consistently oversampled using identical procedures. (4) For the remaining communities, the 1952 sample used variable sampling rates, the 1970 sample was based on selecting every ninth name from the census, and the 1992 sample used computer techniques to make a random selection from the census.

We have found that being a head of household was not related to the prevalence of depression and thus this report draws on all subjects with complete information (Table 1). Subjects gave informed consent to be interviewed. After the first survey, follow-up interviewing of subjects from earlier samples was coordinated with interviewing new subjects. \[25-29\] Because of this, interviewing for the second and third samples took longer than interviewing for the first sample. For the 1970 sample, interviewing extended from the summer of 1969 through 1971, and for the recent sample, from 1992 through 1995. Therefore, current point prevalence should be interpreted as referring to the disorders present at the time of interview.

INTERVIEW METHODS AND DEFINITIONS OF DEPRESSION

Two methods of data gathering and analysis employed herein were developed within the Stirling County Study and have been described elsewhere. \[15,16\] They are named for the diagnostic locus on depression and anxiety (DPAX [DP for depression and AX for anxiety]), with the second (DPAX-2) being a revision of the first (DPAX-1) and with depression being separately analyzed in this report.

The original interview schedule used in the Stirling County Study contained only one question about dysphoria (using the word spirits to represent mood) and only one about functional impairment. According to psychometric principles, the greater the number of questions asked on a given theme, the greater the reliability. \[26\] Thus, for the 1970 phase, several questions were added and they have been maintained in the subsequent issues of the schedule. By the 1990s, positive responses to the question about spirits had declined, while the proportion who responded positively to other questions about dysphoria (“feeling low and hopeless” and “wondering if anything is worthwhile anymore”) increased. This suggested that phrases such as poor spirits or low spirits, like melancholy, were waning as familiar ways of talking about depressed mood. Thus, the DPAX-2 was prepared to make use of the new questions.

These methods employ a concept of depression that approximates the one used in the DSM-III and DSM-IV. \[31-33\] Both the original and revised versions follow an algorithm that takes into account “essential features,” “associated symptoms,” “impairment,” and “duration.”

For essential features, the DPAX-1 requires a response of Sometimes or Rarely to the question “Do you feel in good spirits?” responses that we will describe by the term poor spirits. The DPAX-2 requires a positive response to at least 1 of 3 questions (poor spirits, low and hopeless, or wondering if things are worthwhile).

For associated symptoms, the DPAX-2 needs positive evidence in each of the 3 clusters of symptoms concerning disturbances of appetite, sleep, and energy, but the DPAX-1 allows only 2 clusters to be represented if the subject is frequently in poor spirits. It is in the area of associated symptoms that both versions of the DPAX are limited in congruence to criteria spelled out in DSM-III and DSM-IV.

RESULTS

ESTIMATES OF CURRENT DEPRESSION FOR THREE SAMPLES

The overall prevalence of depression according to the DPAX-1 remained stable at 5.3% in 1952 and 1970 but decreased significantly to 2.9% in 1992 (Table 2). The rates for women were somewhat higher than those for men in each year but the differences were not significant. Age was significantly related to the rates of depression in 1952 and 1970 but not in 1992. However, the patterns of association with age were not consistent over time or by sex.

The DPAX-2 indicated that the rate of depression was 5.3% in 1970 and 5.7% in 1992. In 1970, the prevalence of depression in men and women was not significantly different. In 1992, the rate for women was significantly higher than that for men. Age was significantly related to depression in 1970 but not in 1992.

The DIS results for 1992 indicated that the 1-month rate for MDE was 2.6%, while the rate for dysthymia without MDE was 2.9%; when these rates were combined, the current prevalence of acute and chronic depression was 5.5%. Women had a significantly higher prevalence of the combined disorders than men, but age was not significantly associated with the rate.

AGREEMENT AND DISAGREEMENT BETWEEN THE DPAX-1 AND THE DPAX-2

To explore reasons why the 2 DPAX methods gave different rates in 1992 in contrast to the similarity of rates in 1970, we compared the results as they applied to individual cases. Agreement was better in 1970 than 1992 (Figure 1). To meet the criteria for essential features in the DPAX-2, at least 1 of 3 indicators of dysphoria was...
For functional impairment, both DPAX methods are similar to the model provided in DSM-IV, in which the following criterion was added: “The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.” The DPAX-1 requires a positive response to its one question about “going easy on work,” while the DPAX-2 involves a larger array of questions. Two of them are general and deal with not being able to “get going” and not being “healthy enough to do things,” while the others are linked specifically with depression and concern hospitalization and interference with work, family, and social activities.

Where duration is concerned, the DPAX-2 estimates the time of occurrence and the duration of the episode using supplemental questions, which are asked if positive evidence has been given about the essential features. These questions concern when the symptoms started, whether they still bother the subject, and, if not, when they stopped being bothersome. For the DPAX-1, time of occurrence is based on the period of going easy on work. Both algorithms require a minimum duration of 1 month, but the actual durations have been found to be longer. The average duration of essential symptoms found by the DPAX-2 varied between 10 and 11 years in 1970 and 1992; the average duration of impairment found by the DPAX-1 was about 8 years in each of the 3 surveys. This information does not mean that all aspects of essential features, associated symptoms, and impairment occurred together throughout these durations. Nevertheless, a sizable proportion of the DPAX depressive disorders seemed to be chronic.

The complete module for depression from the DIS, as used in the ECA, was added to the Stirling interview protocol for the 1990s. The purpose was to compare the DPAX findings with those of the ECA and other DIS studies that use DSM-III criteria, especially the other main North American investigation, the Edmonton Psychiatric Epidemiology Study. Major depressive episode (MDE) and dysthymia were used in combination for comparison with the DPAX, since the latter involves both acute and chronic disorders. One-month and lifetime rates of MDE were computed. One-month rates include episodes in existence at the time of interview or within the month preceding the interview. Lifetime rates include all episodes recalled as occurring at any time during the subject’s life up to and including the time of interview. We followed the convention of the ECA in considering dysthymia to be a current disorder.36

STATISTICAL METHODS

The dependent variable for this study is the current prevalence of depression. Three explanatory variables are used: year of study, sex, and age. The main statistical procedure employed was hierarchical modeling, using a form of logistic regression known as log-linear analysis. For analysis focused on between-year comparisons, year of study was investigated first in terms of a model that assessed its association with current depression controlling for sex and age, then by models that assessed the 2-way interactions of year of study with sex and age. The levels of significance were based on maximum likelihood statistics with appropriate degrees of freedom. To facilitate comparison with other investigations, within-year analysis was also carried out. Each of the 3 years was assessed individually for the associations of sex and age with current depression.

To adjust for variable sampling rates, we divided the county into 3 districts (2 for the oversampled and contrasting communities and 1 for the remainder of the county). The rates were standardized by the direct method to the pooled adult populations of the county for 1952, 1970, and 1992.5 The age profile of the population of Stirling County changed in ways similar to that for North America generally because of the increase of births after World War II. Since our goal was a comparison of rates over the years of study, we standardized the rates to the pooled population rather than choosing one year as opposed to another.

### Table 1. Data Completeness, Age, and Sex for the 3 Samples

<table>
<thead>
<tr>
<th>Year of Sample Selection</th>
<th>1952</th>
<th>1970</th>
<th>1992</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. selected</td>
<td>1098</td>
<td>1369</td>
<td>1618</td>
</tr>
<tr>
<td>Complete interviews, No. (%)</td>
<td>1003 (91)</td>
<td>1201 (88)</td>
<td>1396 (86)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age, No. (%)</th>
<th>Female</th>
<th>Male</th>
<th>Female</th>
<th>Male</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;45 y</td>
<td>281 (51.3)</td>
<td>192 (42.1)</td>
<td>271 (44.9)</td>
<td>242 (40.5)</td>
<td>333 (43.9)</td>
<td>280 (43.9)</td>
</tr>
<tr>
<td>45-64 y</td>
<td>175 (32.0)</td>
<td>168 (36.8)</td>
<td>221 (36.6)</td>
<td>234 (39.2)</td>
<td>249 (32.9)</td>
<td>230 (36.0)</td>
</tr>
<tr>
<td>≥65 y</td>
<td>91 (16.6)</td>
<td>96 (21.0)</td>
<td>112 (18.5)</td>
<td>121 (20.2)</td>
<td>176 (23.2)</td>
<td>128 (20.0)</td>
</tr>
<tr>
<td>Total</td>
<td>547 (100)</td>
<td>456 (100)</td>
<td>604 (100)</td>
<td>597 (100)</td>
<td>758 (100)</td>
<td>638 (100)</td>
</tr>
</tbody>
</table>

* In some columns, percentages do not add to 100% because of rounding.
DPAX-1, while the DPAX-2 has multiple indicators of disability. These 2 types of disagreements could be expected to lead to higher prevalence according to the DPAX-2, because it offered more options to describe mood and impairment. However, another type of disagreement constrained such a tendency—the DPAX-2 had more stringent criteria for associated symptoms than the DPAX-1.

The 2 DPAX methods produced similar prevalence rates in 1970 because positive responses about poor spirits had not yet declined seriously and because the number of subjects who did not meet the DPAX-1 criteria for poor spirits was about equal to the number who did not meet the DPAX-2 criteria for associated symptoms. The reason the 2 methods agree less well in 1992 relates primarily to the decreased use of poor spirits and secondarily to declining use of the concept of going easy on work as an indicator of impairment.

This comparison of the 2 DPAX methods suggests that the 1992 results based on the DPAX-1 are probably related to changing vernacular usage rather than to changing rates.

COMPARISON WITH OTHER DIS STUDIES FOR MDE

Comparisons based on using the DIS in 3 North American studies (ECA, Edmonton, and Stirling County) can be made for MDE as a separate disorder rather than in combination with dysthymia, as in the rest of this report. Overall lifetime rates for MDE varied across the studies only from 6.3% to 8.6%, and 1-month rates from 2.2% to 2.6% (Table 4). In all comparisons, the rates for women were about 2 times higher than those for men.

---


<table>
<thead>
<tr>
<th>Method</th>
<th>1952 (n = 1003)</th>
<th>1970 (n = 1201)</th>
<th>1992 (n = 1396)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;45</td>
<td>45-64</td>
<td>≥65</td>
</tr>
<tr>
<td>DPAX-1†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>4.5 (1.2)</td>
<td>6.1 (1.9)</td>
<td>11.0 (3.3)</td>
</tr>
<tr>
<td>Male</td>
<td>2.3 (1.1)</td>
<td>7.0 (2.0)</td>
<td>4.4 (2.1)</td>
</tr>
<tr>
<td>Total</td>
<td>3.4 (0.8)</td>
<td>6.6 (1.4)</td>
<td>7.9 (2.0)</td>
</tr>
<tr>
<td>DPAX-2‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3.3 (1.1)</td>
<td>8.1 (1.9)</td>
<td>8.9 (2.7)</td>
</tr>
<tr>
<td>Male</td>
<td>2.8 (1.1)</td>
<td>9.3 (2.0)</td>
<td>1.9 (1.3)</td>
</tr>
<tr>
<td>Total</td>
<td>3.0 (0.8)</td>
<td>8.7 (1.4)</td>
<td>5.6 (1.6)</td>
</tr>
<tr>
<td>DIS§</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>9.3 (1.7)</td>
<td>7.6 (1.8)</td>
<td>4.8 (1.7)</td>
</tr>
<tr>
<td>Male</td>
<td>3.7 (1.2)</td>
<td>9.3 (2.0)</td>
<td>1.9 (1.3)</td>
</tr>
<tr>
<td>Total</td>
<td>6.4 (1.0)</td>
<td>5.2 (1.0)</td>
<td>3.8 (1.1)</td>
</tr>
</tbody>
</table>

*DPAX-1 and DPAX-2 indicate the original and revised methods of the Stirling County Study method for diagnosing depression (DP) and anxiety (AX); DIS, Diagnostic Interview Schedule. The DPAX rates are point prevalence rates for depression, and the DIS rates combine major depressive episode (1-month prevalence) and dysthymia.

†Between-year comparisons for DPAX-1 indicated that the year of study was not significant for 1952 vs 1970 but was significant for 1970 vs 1992 (\( \chi^2 = 8.85; P = .003 \)). There were no significant interactions. Within-year comparisons indicated that sex was not significant for any year but age was significant in 1952 (\( \chi^2 = 7.08; P = .03 \) and also in 1970 (\( \chi^2 = 8.11; P = .02 \) but not in 1992.

‡Between-year comparisons for DPAX-2 indicated that the year of study was not significant. Sex did not interact significantly with year of study but age did (\( \chi^2 = 7.96; P = .02 \)). Within-year comparisons indicated that age was significant in 1970 (\( \chi^2 = 13.33; P < .001 \)) but not in 1992, while sex was not significant in 1970 but was significant in 1992 (\( \chi^2 = 5.95; P = .02 \).

§Within-year comparisons for DIS indicated that sex was significant (\( \chi^2 = 10.16; P = .001 \)) but age was not.

---

**Figure 1. Levels of agreement between the DPAX-1 and the DPAX-2 in 1970 and 1992. Numbers are numbers of subjects. DPAX-1 and DPAX-2 (DP for depression and AX for anxiety) indicate the original and revised methods for diagnosing depression.**

DPAX-1 and DPAX-2 disagree in 2 main domains: impairment or associated symptoms. The 2 methods agree less well in 1992 relates primarily to declining use of the concept of going easy on work as an indicator of impairment.

This comparison of the 2 DPAX methods suggests that the 1992 results based on the DPAX-1 are probably related to changing vernacular usage rather than to changing rates.
for men. From both the lifetime and 1-month perspectives, younger people had higher prevalence than older people. In the Stirling County and Edmonton studies, the rates for the 2 younger age categories were more uniform than in the ECA study. However, the ratios comparing subjects younger than 45 years with those 65 years or older exhibit 2-fold to 4-fold differences.

TIME, SEX, AND AGE

Based on the evidence that the DPAX-1 findings for 1992 were influenced by vernacular changes, the most appropriate time comparisons from this study involve the DPAX-1 for 1952 and the DPAX-2 for 1970 and 1992, the latter being selected for the last 2 samples because its coverage of impairment is more congruent with the intent of the disability criterion of DSM-IV and because its requirement for 3 associated symptoms is a closer approximation of the 4 needed according to both DSM-III and DSM-IV. Using this selection of findings, it seemed worthwhile to compare current prevalence using a single threshold for age (<45 years vs ≥45 years). This threshold was chosen because the ECA findings of higher prevalence among those younger than 45 years have drawn considerable attention and influenced the design of the National Comorbidity Survey in terms of its focus on younger persons. Under these circumstances, the DPAX rates of depression in the 2 early samples were quite similar for women and men, and prevalence increased with age for both women and men (Figure 2). In 1992, the DPAX and DIS rates for women were similar and different from those in the earlier years. Both methods emphasized that the highest prevalence of depression was among younger women. The rates for younger men were similar across time and across method. The DPAX and DIS rates for older men differed, however, in that the DPAX showed the prevalence of depression rising with age, while it declined somewhat according to the DIS.

The conclusion we draw from this longitudinal analysis is that the current prevalence of depression in Stirling County remained steady at about 5% among successive samples of adults over the period from 1952 to 1992 and that these depressions tended to be chronic. At the end point of the study, the current prevalence of acute and chronic depression based on the DIS method was also about 5%.

At the same time, this study illustrates the problems posed for time-trend analysis by changes in the language people use to describe their moods and by the evolution of diagnostic criteria. There are 3 ways in which time can influence prevalence: show an increase, show a decrease, or maintain a stable rate. Each of these possibilities was, to some extent, evident. The DPAX-1 indicated a decrease by 1992, the DPAX-2 indicated stable...
prevalence, and the use of recalled lifetime evidence based on the DIS gave information of the same type that in other studies has been interpreted as showing an increase.

Our analysis of the DPAX-1 results indicated that the decrease in prevalence was influenced by changes in the vernacular and should therefore be discounted as showing a real decrease in prevalence. Where diagnostic standards are concerned, the methods used throughout this study share with contemporary approaches the concept that a diagnosis involves essential features, associated symptoms, duration, and impairment. At the level of several of the specific criteria, however, the DPAX and DIS methods differ, pointing to a dilemma for analyzing time trends. Use of the specific criteria of DSM-III, as with the DIS, depends on reconstructing the past by asking people to recall their psychiatric histories; such recollections may involve inaccuracies. On the other hand, data that do not require long-term memory, as in the use of current DPAX depression in samples dating back to 1952, cannot take into account all of the specific criteria that became standardized in DSM-III.

A question can be raised as to whether there is any common ground between the findings about a steady current prevalence of depression and findings that indicate increasing lifetime rates of depression. A stable current prevalence rate is not necessarily incompatible with increasing lifetime rates. If most of the disorders registered in the lifetime rates were short-lived, either because of their nature or because of effective treatment, their increase would have only small influence on current prevalence. Thus, at least theoretically, rising lifetime rates of acute depression and steady prevalence of chronic depression can occur together.

Increasingly, it is recognized that depression is an etiologically complex disorder that often runs a recurrent or chronic course. In reporting the Stirling County findings comparing 1952 and 1970, we suggested that stable prevalence is more realistic than marked change for a type of disorder that exhibits such features. It seems unlikely, for example, that genetic and social circumstances would combine in a coordinated way during the same period to foster a pronounced increase or decrease. If changes were found to occur in a circumcribed historical period, it is likely that they would reflect social influences, but even where social experiences are concerned, there are cross-currents of advantages and disadvantages in a population at large that may give protection for some and have noxious effects for others. Thus, small changes in the distribution of depression may be more credible than wide fluctuations.

In considering these possibilities, the generalizability of the information from this study is an important issue. If the environment of the study area is unique and homogeneous, the findings would not have as much general relevance as if the region were typical of many others. Like most rural areas, this one has not grown much in population size, the way cities have, but its population is heterogeneous and its history has been molded by the same forces commonly observed throughout North America. In terms of psychiatric features, the DIS information from Stirling County in the present decade indicates that the prevalence of MDE is similar to what has been found elsewhere. This raises the possibility that other places might have appeared similar to Stirling County if the DPAX methods had been used. The population studies of similar vintage, such as the Lundby and Midtown Manhattan investigations, did not use structured diagnostic methods, thus making comparison problematic. Many of the other earlier studies used short inventories of symptoms with cases identified by a cutoff point rather than diagnostically oriented interviews.

While the methods used in this study (the DPAX and the DIS) show advances over the use of symptom inventories, both have limitations. It has been suggested that when the available methods for tracing trends are imperfect, it is useful to “pursue multiple pathways and look for agreements across them” (L. E. Robins, PhD, letter, March 3, 1998). Since 1980 most studies have employed either the DIS or the Composite International Diagnostic Interview (CIDI). The findings have been consistent in regard to 2 points. One is that the prevalence of MDE is about twice as high among women as men, and the other is that both lifetime and recent prevalence rates are higher among younger than older persons. The results from the 1952 and 1970 samples in Stirling County were not congruent with these DIS and CIDI results. After 1970, however, the DPAX methods indicated that, within an overall stable rate, women were about twice as likely to report depression as men and that the burden of prevalence was moving toward younger women. In these terms, then, the DIS and DPAX methods provide mutual corroboration, although the DPAX findings point to the fact that the associations seen with consistency in DIS and CIDI studies may be of fairly recent origin.

This study suggests that interpretations about a general increase in depression should be tempered or modified in favor of a redistribution of prevalence by sex and age. The higher prevalence of depression among women younger than 45 years in the recent survey in Stirling County is especially worthy of attention. This change refers to a specific time span and a specific segment of the population. Therefore, it may be possible to develop testable hypotheses for investigation that would throw light on the origins of change, especially as these refer to the portion of prevalence that may be under the influence of social and historical circumstances.

Accepted for publication April 29, 1999.

This work was supported by grant MH39576 from the National Institute of Mental Health, Rockville, Md (Dr Murphy).

A preliminary version of this article was presented at grand rounds, Department of Psychiatry, Dalhousie University Faculty of Medicine, April 15, 1998, Halifax, Nova Scotia.

We thank Patricia Merritt, Barbara Burns, and Ellen Krystofik for field management and data processing; Ellen Krystofik and Barbara Burns for manuscript preparation; and Stephen C. Newman, MD, and Roger C. Bland, MB, for providing the lifetime and 1-month rates from the Edmonton study in the age categories used in this article.

Reprints: Jane M. Murphy, PhD, Department of Psychiatry, Room 9155, Massachusetts General Hospital, 149 13th St, Charlestown, MA 02129-2000 (e-mail: murphy@al.mgh.harvard.edu).