Acta Psychiatr Scand 2004: 109: 355–375
Printed in UK. All rights reserved

ACTA PSYCHIATRICA SCANDINAVICA

Anxiety and depression: a 40-year perspective on relationships regarding prevalence, distribution, and comorbidity

Murphy JM, Horton NJ, Laird NM, Monson RR, Sobol AM, Leighton AH. Anxiety and depression: a 40-year perspective on relationships regarding prevalence, distribution, and comorbidity.
Acta Psychiatr Scand 2004: 109: 355–375. © Blackwell Munksgaard 2004.

Objective: Building on a report about the prevalence of depression over time, this paper examines historical trends regarding anxiety in terms of its prevalence, its distribution by age and gender, and its comorbidity with depression. Methods for conducting such time trend analysis are reviewed.
Method: Representative samples of adults were selected and interviewed in 1952, 1970, and 1992. Logistic regressions were used for statistical analysis.
Results: Although twice as common as depression, the prevalence of anxiety was equally stable. Anxiety was consistently and significantly more characteristic of women than men. A re-distribution of rates in 1992 indicated that depression but not anxiety had significantly increased among younger women ($P = 0.03$). Throughout the study, approximately half of the cases of anxiety also suffered depression.
Conclusion: The relationships between anxiety and depression remained similar over time with the exception that depression came to resemble anxiety as a disorder to which women were significantly more vulnerable than men. Social and historical factors should be investigated to assess their relevance to this change.

J. M. Murphy[1,2,3], N. J. Horton[4], N. M. Laird[5], R. R. Monson[3], A. M. Sobol[1], A. H. Leighton[6]

[1]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, [2]Department of Psychiatry, Harvard Medical School, Boston, MA, [3]Department of Epidemiology, Harvard School of Public Health, Boston, MA, [4]Department of Mathematics, Smith College, Northampton, MA, [5]Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA and [6]Department of Psychiatry, Faculty of Medicine, Dalhousie University, Halifax, Nova Scotia, Canada

Key words: anxiety; depression; comorbidity; epidemiology; time factors

Jane M. Murphy, Department of Psychiatry, Massachusetts General Hospital, Room 10028, 149 13th Street, Charlestown, MA 02129-2000, USA
E-mail: murphy.jane@mgh.harvard.edu

Accepted for publication December 9, 2003

Introduction

The concept of 'an age' has been employed to depict how one generation may differ from another in regard to the prevalence and incidence of psychiatric disorders. Pointing to increased public awareness of depression during the 1970s, Klerman speculated that 'in contrast to the "age of anxiety" that followed World War II, we may now be entering an "age of melancholy" precipitated in part by recent global events and consequent "doomsday" prophecies concerning nuclear warfare, overpopulation, and ecological destruction' (1, 2). For the title of a report about increasing depression in the Lundby Study, Hagnell and co-workers used 'Are We Entering an Age of Melancholy?' (3). Employing 'The Age of Anxiety?' as a title, Twenge focused on anxiety and neuroticism among young people suggesting that 'worry about safety, social acceptance, and job security' has increased (4).

Evidence supporting these views has come from different perspectives. Klerman drew on two perspectives. One was *treatment* records, which suggested that by 1970 psychiatrists were beginning to see depressed patients who were younger, less severely ill, and more commonly neurotic than psychotic (5). The other perspective was epidemiologic studies of psychiatric disorders in general populations, as illustrated in the US Epidemiological Catchment Area (ECA) Program (6). In the latter, *lifetime prevalence* rates indicated to Klerman and others that depression was increasing in the form of a 'birth cohort effect' that was exhibited by depression occurring at younger ages and at higher rates among those born recently compared with those born earlier (7–10). Birth after the World War II has been taken as an approximate marker for when the 'birth cohort effect' became noticeable.

The lifetime prevalence rates reported in the ECA can be thought of as *retrospective incidence*.

They were based on what subjects could recall about their first experience with depression. In contrast, the Lundby Study reported an increase of depression by using *prospective incidence*. After an initial assessment, subjects were followed over time to identify first occurrences of depression among those at risk.

Each of the above perspectives dealt with *diagnosed disorders* by which we mean that the case was diagnosed by a psychiatrist, as in the Lundby Study and in treatment studies, or that, as in the case of the ECA, diagnostic interview methods were used. Twenge, however, focused on *symptom levels* and utilized meta-analysis of several studies. This indicated that the mean levels of anxious symptoms were higher among those born more recently than those born earlier.

Thus from these different perspectives, interest has been expressed about whether one of these two common psychiatric disorders is replacing the other and whether one or both are increasing. Changes of these types, if well documented, are important because they may give clues about etiology that can be used to guide future research.

One of the main aims of this paper is to examine questions about changing rates using still another perspective, that of the *current prevalence* of both depression and anxiety based on information from three different samples of adults that were selected from a general population in 1952, 1970, and 1992. In this, we build on and extend an earlier article about the current prevalence of depression (11). While symptom levels are involved in the assessment of both disorders, the evidence presented here concerns *diagnosed disorders* in the sense that diagnostic algorithms were used to identify syndromes, to apply criteria for their completeness, and to rate both duration and impairment.

By focusing on current prevalence, we show the proportions of subjects who met the criteria for diagnosis at the time of interview. Thus, current prevalence does not depend on recalling when the disorder first emerged. Unlike incidence, however, a current prevalence rate is more influenced by chronic than by acute disorders because it enumerates those conditions that persist into the present.

In previous work on depression, we found that the overall prevalence of depression, enumerating together both chronic and acute disorders, remained steady at about 5% in each of the three samples (11). In the samples of 1952 and 1970, women were somewhat but not significantly more prone to depression than men. Older people, however, were significantly more vulnerable than younger. In 1992, a re-distribution of prevalence was observed. The rate among women who were younger than 45 years was about twice what it had been for women of similar ages in the two earlier samples. This change was counterbalanced by small decreases in the rates for older men and women so that the overall prevalence remained stable up to and including 1992. Due to the increase among younger women, the difference in rates comparing men and women was significant in the 1992 sample and the ratio was close to the 2 to 1 relationship that has been found in many recent studies (12, 13).

We interpreted the evidence about depression among younger women in the 1992 sample as supporting the notion of a 'birth cohort effect'. While recognizing the difficulty of disentangling 'age', 'period', and 'cohort' effects, our view is based on the following: (i) all of the women under 45 years of age in the 1992 sample had been born after the World War II while none of the women in the 1952 sample and only a few in the 1970 sample had been born that recently; (ii) the onsets of the depressions experienced by these recently born women did not cluster around a given year or a given period of any 5–10 years but rather were highly variable in terms of time of occurrence, thus making it unlikely that onset was related to an event at a given period in historical time; and (iii) our study of the incidence of depression, based on follow-up of the 1952 and 1970 samples, indicated that women in these cohorts did not exhibit an increase of depression as they grew older although they had lived through the same years following World War II as those born after the war, thus arguing against both a period and an age effect (14).

In addition to investigating the relationships between depression and anxiety in terms of prevalence and the distribution of prevalence by gender and age, we also assess the degree to which the two disorders co-occurred. Thus, another main aim in this paper is to investigate whether there were changes over time in the affiliation of the two disorders. In recent years, considerable attention has been drawn to high levels of comorbidity among psychiatric disorders (15–19). The association between depression and anxiety, especially generalized anxiety, has aroused not only interest but also controversy. Some investigators emphasize that generalized anxiety and depression are distinct disorders and that, when they co-occur, they illustrate comorbidity (20–23). Others stress the difficulty of distinguishing between them (24–26). Still others suggest that generalized anxiety is best thought of as 'a prodromal, incomplete, or residual manifestation

of other psychiatric disorders' (27). These questions about the overlap between the two syndromes are pertinent to interpreting the nature of their relationship.

While there is interest in whether the passage of time has changed the relationships explored in this paper, there are several difficulties posed by the very fact that time goes by. Diagnostic criteria and the methods of applying them may change as may also the levels of public and private recognition of information to which criteria need to be applied (28). Our experiences over the 40-year span of the study suggest that an important problem has to do with changes in the words people use to describe psychiatric phenomena. Because of this, we developed a strategy that accounts for vernacular changes and employed it in our analysis of the prevalence of depression. In this paper, we test the appropriateness of the strategy for analyzing anxiety.

Aims of the study

1. A *methodologic* aim is to evaluate a strategy for studying time trends in anxiety by comparing the results to those for depression.
2. An aim regarding *comorbidity* is to investigate the overlap between the syndromes to see if changes occurred over time and if the evidence contributes to an understanding of whether the relationship is one between two distinct disorders or, alternatively, of anxiety being an 'incomplete' disorder.
3. An aim relevant to *prevalence* is to see whether anxiety increased or decreased while depression remained stable.
4. An aim relevant to *distribution* is to find out if there were changes over time in the way anxiety was associated with age and gender that paralleled the 'birth cohort effect' observed for depression among women.

Material and methods

Site

Data for this report have been gathered in an area of Canada that was given the pseudonym of 'Stirling County' to protect identity. The site is located on the eastern side of the Gulf of Maine, has a population that has consistently numbered about 20 000, and was typically rural at the beginning of the investigation (29–31). Over the years, major changes have occurred. Primary industries have declined while service and information industries have grown. Women have become a large part of the labor force, educational levels have risen, health care has undergone a transition from private practice to national insurance, the standard of living has gone up, the media and travel have widened knowledge, racial tension has increased as has also the use of drugs and attendant criminal activity. Family size has decreased and the bonds of marriage, religion, and community have weakened. While not itself urbanized in the sense of being densely populated, the area has undergone most of the changes that characterize industrial societies generally (11, 32, 33).

In terms of psychiatric characteristics, we have compared the population of Stirling County to other North American populations using a common methodology. For the collection of data carried out in the 1990s, we used not only our own interview schedule but also the depression portions of the Diagnostic Interview Schedule (DIS), the instrument that had been used in the ECA (34). This means it was possible to compare our findings about Major Depressive Episodes, as defined in the third revision of the American Diagnostic and Statistical Manual (DSM-III), with those in the ECA and the Edmonton Psychiatric Epidemiologic Study (35–39). The 1-month rates were nearly the same (Stirling 2.6%, ECA 2.2% and Edmonton 2.3%) and the lifetime rates were very similar (7.9, 6.3, and 8.6%) (11).

Thus, we interpret that both the psychiatric and social features of the population are like those found in many other places, especially those in North America and other industrialized areas.

Design of study and samples

Two longitudinal approaches have been employed: (i) successive samples of the population have been surveyed in order to estimate trends in prevalence (11, 32); and (ii) follow-up with individuals of previous samples has been conducted in order to investigate incidence, precursors, course, and outcome (14, 40–43). This dual approach was employed so as to provide, across a period of 40 years, follow-up information on the population as a whole as well as on the individuals who compose it. In this article, we focus on following the population by means of repeated cross-sectional surveys. Following individuals over time is, however, a more common strategy in psychiatric epidemiology. Aside from the Stirling Study, another example of the design of successive samples of a population is the comparison of the original sample of the US National Comorbidity Survey carried out in 1992 with a new sample in 2002 (44).

In this report, the three samples of adults that provided information for estimates of prevalence consisted of 1003 persons for 1952, 1201 for 1970, and 1396 for 1992 (Table 1). The procedures of sampling have involved the creation and updating of enlarged aerial maps showing residential houses and a research census of the adults living in them (11, 30, 32). Of the 100 communities, five have consistently been over-sampled because they illustrate marked contrast in regard to social and economic resources. The sampling ratio changed over time, primarily to achieve progressively larger samples.

For the 1952 sample, only heads of households ('main breadwinner' for men and 'main preparator of food' for women) were selected but subsequently the sampling frame focused on all adults 18 years of age and older. The change to all adults was made for purposes of comparability to other studies. For time comparisons within the Stirling Study, we investigated the effect of household position and found that it had minimal influence on rates of psychiatric disorders (11, 32). Thus the two recent samples as presented here were not restricted to household heads.

In 1952, households were preselected as every *n*th house located on the maps; and subjects were selected by interviewers alternating between male and female heads. In 1970, subjects were preselected by hand from an updated census. In 1992, subjects were preselected by computer from an updated census. Examination of the samples has indicated that, with the exception of a higher completion rate among women compared with men, the demographic features of the population as a whole are proportionately represented.

In this paper, we use the age categories of <45 and $45+$ because they give large enough numbers to provide stable rates and because recent epidemiologic studies suggest that this division is relevant to some of the most important differences between younger and older subjects (6).

Collection of data

Subjects have been interviewed in their homes by lay interviewers trained in the administration of a structured interview schedule. The subjects were informed about the purpose of the research and gave consent to be interviewed. The procedures for informed consent have been approved by government and academic committees for ethical research in both the US and Canada.

The interview schedule concerns general health, depression, anxiety, and impairment in everyday functioning. It also covers a number of variables about social conditions and life experiences. The core of the schedule has not been altered over the years.

Because the information gathered by means of interviews with subjects was limited to depression and anxiety, interviews with the general physicians of the county have been conducted so as to compile a more adequate base for estimating the prevalence of all types of psychiatric disorders (45). These interviews were carried out by psychiatrists and they focused on descriptions of symptoms that are of concern in a psychiatric work-up as well as the medical histories of the sample members. In this report, we deal almost exclusively with data about depression and anxiety as gathered in the interviews with subjects.

Depression and anxiety

In 1948 when we began preparing for the Stirling County Study, most of what was known about the

Table 1. Data-completeness, gender and age for three samples*

	Year of sample selection					
	1952		1970		1992	
	Female	Male	Female	Male	Female	Male
Age (years)						
<45, *n* (%)	284 (51.9)	197 (43.2)	276 (45.7)	247 (41.4)	333 (43.9)	280 (43.9)
≥45, *n* (%)	263 (48.1)	259 (56.8)	328 (54.3)	350 (58.6)	425 (56.1)	358 (56.1)
Total	547 (100)	456 (100)	604 (100)	597 (100)	758 (100)	638 (100)
Number selected	1098		1369		1618	
Complete interviews	1003		1201		1396	
Completion rates	91%		88%		86%	

* All members of the 1952 and 1970 samples were followed up at later points in time. A few gave slightly different information about their ages in subsequent interviews. For those who died, information about age on the death certificates did not always agree perfectly with what was given in the interviews. Such discrepancies have been resolved. Although this report does not include any follow-up information, the resolution of discrepancies means that the ages shown in this table differ somewhat from those given in previous publications. The resolutions had very little effect on the findings, usually no more than a small portion of a percentage point. There were no changes in the statistical findings due to these resolutions.

epidemiology of psychiatric disorders was based on treatment records and studies in clinic settings where diagnoses were provided by psychiatrists. While the World War II had emphasized the importance of learning about the amount and kinds of psychiatric disorders found in a general population, there was little experience in conducting psychiatric epidemiologic investigations among community residents. One model for such investigation was provided in the Swedish Lundby Study that had been started in 1947. It involved having a group of psychiatrists trained in the same university examine all members of a population of 2500 including not only adults but also infants and children (46–48). In planning the Stirling County Study with its population of 20 000, we realized that cost and logistics would make a psychiatric examination of each member of the population impossible. Further, the training of psychiatrists in North America was sufficiently diverse that the examinations would not necessarily result in comparable and reliable diagnosis about each subject. Nor was the field of child psychiatry sufficiently well developed at that time to make it practical to include those younger than 18.

The World War II had, however, provided reason for researchers in the US and UK to develop systematic inventories of questions that could be administered to large numbers of service personnel, albeit mainly male personnel (49–51). It seemed wise, therefore, to draw on the war experiences in this regard, and we conducted an advance study concerned with choosing appropriate questions for a general population that would include men and women of a broad age range (52).

The questions about depression and anxiety that came to constitute the core of our interview schedule were selected empirically from a large battery of questions mainly taken from the inventories developed during World War II. These were administered to a group of 78 patients diagnosed as suffering from what was then called a 'psychoneurotic disorder' and a sample of 559 community subjects selected as essentially healthy. The community work was carried out in a geographic area that was similar to although separate from Stirling County so as to avoid the need to study two fairly large samples at the same time in the main study site. The selected questions were those that discriminated between the patients and community residents at statistically significant levels. As part of this preparatory study, an evaluation of validity was carried out by a psychiatrist who was blind to the question responses and who conducted face-to-face assessments of a subsample of community subjects in order to record whether he believed the subject to be a 'psychiatric case' or not. He was not asked to assign a diagnosis but rather simply to indicate if he thought that the person suffered from some kind of psychiatric disorder. The results can be summarized as indicating that an appropriate threshold score gave 83% sensitivity and 76% specificity (Table 2).

A point of interest is that the inventories from which our questions were selected did not include a direct question about depressed mood although they covered most of the associated symptoms of depression. The absence of a question about lowered mood seems to have followed from the fact that, compared with anxiety, depression was conceived to be less common, more serious, and often psychotic in its features. Nevertheless, we believed it essential to have a question about dysphoria, and we therefore added to our schedule a question about being in 'poor spirits'.

For the analysis of the physician and subject information from the 1952 sample, as given in the first reports of the Stirling Study, the responses to these questions were read by psychiatrists who provided case identification and diagnosis (31). Recognizing the importance of assessing the data in a systematic and reliable way, a procedure was developed that can be called a human psychiatric evaluation. The cardinal feature of this procedure was for the psychiatrists to set aside theoretical formulations and to focus on the symptomatic patterning of the phenomena reported by the subjects. To this end, the psychiatrists' decisions were made according to descriptions of the syndromes and guidelines that required a rating of judgmental confidence (in the sense of 'doubt' to 'certainty') that a psychiatric syndrome was manifest in the interview record. Ratings for duration and impairment were also required. At least two psychiatrists independently read each protocol and then met to achieve a series of ratings representing consensus. Inter-rater reliability varied according to diagnosis but in overall terms was quite high. Nevertheless, the consensus ratings were used to report findings (31).

In the mid-1960s, we carried out another investigation of validity, this time in Stirling County, where a subsample was visited in their homes by psychiatrists (53–55). The psychiatrists conducted clinical interviews but were trained to gather the data so that they could record the results in terms of confidence that a psychiatric syndrome was in evidence, duration, and impairment. The survey data about the subsample included the subject's responses to the interview as well as reports from usually two general physicians about the subject. Again the task of the psychiatrists was to make a

Table 2. Evolution of validity investigations in the Stirling County Study

	Standard for assessing validity		Stirling Study method tested for validity		Results of validity testing		
Study year (no. of subjects)	Types of interviewers	Types of interview	Source of information	Mode of analysis	Decision category	Sensitivity (%)	Specificity (%)
1951 (64)	Psychiatrist	Unstructured	Self-report	Threshold score*	Psychiatric case vs. not case	83	76
1964 (120)	Psychiatrists	Semi-structured	Self-report and general physician report	Human psychiatric evaluation†	Psychiatric case vs. not case	93	83
					Impaired case vs. not impaired	69	91
1995 (139)	Psychiatrists and clinical psychologists	Structured clinical interview for DSM-III-R (SCID)	Self-report	Computerized diagnostic algorithms‡			
					Depression and/or anxiety vs. not		
				DPAX-1		51	81
				DPAX-2		70	76
				DIS		67	88
					Anxiety vs. not		
				DPAX-1		52	74
				DPAX-2		69	64
				DIS		74	74
					Depression vs. not		
				DPAX-1		19	89
				DPAX-2		31	89
				DIS		47	95

* A threshold score was applied to a frequency-weighted count of the symptoms described in the questions selected as part of the advance study that was carried out in order to prepare the questions to be asked in the Stirling County Study. The psychiatrist whose judgments served as the standard for validity recorded his decisions in categories for psychiatric cases vs. not psychiatric cases with a middle category being for subjects about whom he was uncertain. In the original publication, the results were not presented as sensitivity and specificity (52). If the doubtful cases had been counted as not psychiatric cases, sensitivity would have been perfect but specificity would have been low. We and others who have reviewed this study suggest that the fairest analysis gives the values shown here.

† The Human Psychiatric Evaluation was a systematic procedure developed in the Stirling Study by which two psychiatrists independently read a protocol that combined the responses by a given subject to the questions in the self-report interviews and the descriptive reports given by general physicians about that subject. The product of the procedure was a subsequent joint assessment reflecting the agreement of the psychiatrists about psychiatric caseness and level of impairment in everyday functioning. Although the psychiatrists provided information about the diagnostic nature of the case, the standard they gave for testing validity was limited to the decision of whether the subject was a psychiatric case or not using categories representing levels of certainty and whether the subject was impaired in functioning or not using categories for level of impairment. In the original publication, the results were not presented as sensitivity and specificity (53). Here the results are shown using the category thresholds for identifying cases counted in prevalence rates.

‡ The Computerized Algorithms pertain to a procedure developed in the Stirling Study and named Depression and Anxiety (DPAX) that provides diagnostic information based on the structured interviews with subjects (58). Two versions of the procedure are used (DPAX-1 stands for the original used throughout the study and DPAX-2 for an updated version used from 1970 onwards). DIS stands for the Diagnostic Interview Schedule developed by the US National Institute of Mental Health and used in the 1990s phase of the Stirling Study.

Table 3. Components and algorithms used in two versions of DPAX

Components	Depression		Anxiety	
	DPAX-1*	DPAX-2*	DPAX-1*	DPAX-2*
Essential features	Poor spirits	Poor spirits	Frightening dreams	Frightening dreams
		Low/hopeless	Palpitations	Frightened not know why
		Wonder things worthwhile	Trembling	Nervousness
Associated symptoms†	Trouble sleeping	Trouble sleeping	Sweating	Sweating
		Tired mornings	Paraesthesia	Paraesthesia
	Loss of appetite	Loss of appetite	Cold sweats	Cold sweats
	Food tasteless	Food tasteless	Sick headaches	Sick headaches
	Many ailments	Many ailments	Upset stomach	Upset stomach
	Tired mornings	Weak all over	Bad taste in mouth	Bad taste in mouth
		Things do not turn out		Palpitations
				Trembling
Orienting feature‡	Apprehension about nervous breakdown	Apprehension about nervous breakdown	Apprehension about nervous breakdown	Apprehension about nervous breakdown
Impairment in everyday functioning§	Going easy on work	Cannot get going	Going easy on work	Cannot get going
		Not healthy to do things		Not healthy to do things
Duration	At least 1 month	At least 1 month	At least 1 month	At least 1 month

* DPAX-1 and DPAX-2: original and revised versions of a diagnostic algorithm named Depression and Anxiety (DPAX).

† The associated symptoms for depression reflect disturbances of sleep, appetite, and energy as well as an extra symptom in DPAX-2 for dejection related to having things not turn out as hoped. For DPAX-1, 'tired mornings' was thought of as related to energy disturbance but in DPAX-2, for which there were extra questions, 'tired mornings' was thought of as related to sleep disturbance.

‡ Apprehension about having a nervous breakdown played an important role in identifying subjects as having a depression or anxiety disorder. It did not, however, help differentiate between the two disorders.

§ For DPAX-2, supplements were asked of those subjects who responded positively to the essential feature questions. The supplements included questions about impairment associated with the disorder whose essential feature led to the supplement. These questions concerned staying away from work, staying in bed, going to the hospital, experiencing problems in the family or in social activities. All the questions shown in the table were asked of all subjects.

judgment about whether the subject suffered from a psychiatric disorder (was a 'psychiatric case') rather than determining what kind of disorder was manifest (53). In addition, however, the psychiatrists needed to make a decision as to whether or not the person was impaired in everyday functioning. The results indicated that the survey data had very good values for 'psychiatric caseness' but somewhat less adequate values for impairment (Table 2). The latter finding emphasized that we needed to improve the types of questions asked about how the person was getting along in family, work, and community settings.

Subsequently a computerized algorithm was designed for making the same kinds of diagnoses based only on the responses given by subjects in the face-to-face interviews (56). The algorithm was named Depression and Anxiety (DPAX) and it laid out steps for identifying the 'essential features' of these disorders, monitoring the completeness of the syndromes as indicated in 'associated symptoms', establishing that the syndromes persisted over a minimum duration, and assessing the degree to which they interfered with everyday functioning. The task of reproducing the psychiatrists' judgments made it clear that not only had the psychiatrists given special weight to the 'essential features' that distinguish the sadness of depression from the fearfulness of anxiety but also to positive responses to a question about feeling that one might have a 'nervous breakdown'. Responses to this question did not distinguish between depression and anxiety, although they played a somewhat stronger role in the replication of anxiety than depression. The computer program reproduced the psychiatrists' ratings with sensitivity of 92% and specificity of 98%. This level of replication was possible due to the explicit guidelines employed by the psychiatrists and to the use of the consensus ratings as the standard. It is to be emphasized, however, that these sensitivity and specificity values do not reflect validity. They indicate, rather, that the computer program reliably reproduced the psychiatrists' ratings.

We suggested that the human and computerized procedures for providing diagnoses in our study are similar to the diagnostic algorithms spelled out in DSM-III (57). As time went by, however, it became clear that two versions of the DPAX algorithm were needed (Table 3). We call the original version DPAX-1 and the revised version DPAX-2 (58).

The main reason for developing the revised version was to take account of secular drift over time in the meaning of words and the frequency with which they are used. The longer the duration of a period of observation, the greater the likelihood that secular changes are apt to occur. In the

nineteenth century, for example, the idiom, 'feeling spleeny', based on the humoral view that melancholy was associated with the spleen, referred to what today would be described as a symptom of depression (59). If interviews carried out at that time had failed to use the 'spleeny' idiom, depression might have gone undetected. Today, most people would probably not know what 'feeling spleeny' means. The idiomatic expressions used at the present time are different as illustrated in the common use of color, as in 'feeling blue' for sadness. Further, it is very probable that still different idioms will be used in the future, as suggested in the fact that 'feeling shitty' seems to be becoming increasingly common as a way of talking about an emotional state that for some people constitutes an important feature of depression. In epidemiologic surveys, it is important to ask questions using words and idioms that are understood by a broad cross-section of the population. We suggest, therefore, that longitudinal research needs to be sensitive to such vernacular changes while still maintaining as much consistency of focus on the phenomena as possible.

A related reason for having two versions of DPAX concerns the advantage of providing multiple words and idioms for a given type of mood or emotion. For the 'essential feature' step regarding depression in DPAX-1, only one question was available, and it was the one cast in terms of 'poor spirits'. In addition, the 'impairment' step for both depression and anxiety was based on only one question that used the idiom of 'going easy' in reference to one's everyday work. After the first survey in 1952, we believed it advisable to expand the number of questions asked about depressed mood, anxious affect, and impairment in functioning. Thus we added extra questions to the protocol, placing them after the basic core of the interview schedule.

For the concept of dysphoria, the added questions pertained to 'feeling low and hopeless', 'wondering if anything is worthwhile any more', and 'feeling that things do not turn out well'. The questions added for the 'essential feature' of anxiety were 'feeling nervous' and 'being frightened without knowing why'. With the latter, it was possible to give greater emphasis to the clinical importance of fearfulness and nervousness in the concept of anxiety than pertained in the original procedures. After these questions, supplements were added for those subjects who gave positive evidence about the 'essential features'. The supplements allowed more specific dating of onset and duration and covered impairment in family and social roles as well as work.

In order to construct a computer algorithm for the protocol that included the new questions, we needed human psychiatric evaluations based on the revised protocol that included the core questions as well as the additional ones. A resource for this was a companion study we had carried out during the mid-1960s in New York City (60). The expanded protocol had been used in this study, and for its evaluation a new group of psychiatrists was trained by the psychiatrist who had played a central role in the design and implementation of the procedure for the 1952 study in Stirling County and for the validity study in the mid-1960s. The same syndrome descriptions were used and the same guidelines were employed for rating confidence about the presence of a psychiatric syndrome, duration, and impairment. As in the earlier work, consensus ratings provided by two independent raters served as the standard for replication. The algorithm for DPAX-2 was found to have 89% sensitivity and 91% specificity for reproducing that standard (57).

DPAX-1 can be applied to information gathered throughout all years of the study; and DPAX-2 can be applied to data collected from 1970 onward. In reporting on depression, we showed that prevalence according to the original version (DPAX-1) remained steady from 1952 to 1970 and then declined significantly in 1992 (11). Investigating this counter-intuitive result, we found that positive responses to the one question about dysphoria in DPAX-1 ('poor spirits') declined. However, responses to the additional questions about dysphoria that were available from 1970 onward increased.

There is considerable evidence that the idea of 'one's spirits being low' was frequently used to describe dysphoria in the past (61). It was also widely employed by psychiatrists when conducting a mental status examination during the period of the 1930s through the 1960s (62, 63). However, the descriptions of symptoms in the recent diagnostic manuals and the newer interview schedules, such as the DIS, the Composite International Diagnostic Interview (CIDI), and the Clinical Interview Schedule-Revised, do not employ 'low spirits' (34, 35, 64–68). Rather they use words and idioms such as feeling 'blue', 'hopeless', 'miserable', 'down in the dumps', and 'blah'. They also use the word 'depressed' as its meaning in reference to mood has now become well known.

We concluded, therefore, that 'poor spirits', like 'melancholy' is passing out of the common vocabulary. Further, we suggested that the decline in prevalence for 1992 according to DPAX-1 was not an accurate picture of what was occurring about

depression as a clinical entity, but rather was a reflection of dependence on an old-fashioned idiom for the central concept of depression (11).

As part of the collection of data conducted in the 1990s, another assessment of validity was made (69). This time a subsample was interviewed by psychiatrists and clinical psychologists who administered the Structured Clinical Interview for Diagnosis (DSM-IIIR) (SCID) (70). Because we had also administered the portions of the DIS that were similar to what was covered in DPAX (major depression, dysthymia for depression and panic and generalized anxiety for anxiety), it was possible to include the DIS in the evaluation of validity. In this study, the task of the clinicians in administering the SCID was different from what pertained in our earlier validity investigations. The focus was on whether or not the subject met criteria for a diagnosis rather than whether the subject should be thought of as a 'psychiatric case'.

These recent validity results were not as good as in our earlier investigations (Table 2). They indicated that agreement was better when the depression and anxiety diagnoses were aggregated together than when each diagnosis was considered separately. Further, agreement was better for anxiety than for depression. Across assessments, agreement was consistently best for the DIS, followed next for DPAX-2, and lowest for DPAX-1. Bearing in mind that the SCID interviews were conducted in the 1990s when we had already become alert to the problem posed for DPAX-1 by dependence on the 'poor spirits' question, it is to be noted that the lowest of the sensitivity values pertained for SCID and DPAX-1 regarding depression, thus giving another bit of evidence that 'poor spirits' was no longer doing a good job of representing dysphoria.

Based on evidence that some of the DPAX questions appeared to have become old-fashioned, we proposed that the most adequate analysis of time trends for our study would combine the two versions of DPAX by utilizing DPAX-1 for 1952 and DPAX-2 for 1970 and 1992. We call this model DPAX122 where the numbers stand for the three study years and for the change from DPAX-1 to DPAX-2 in 1970 and contrast it to the two models based on using only one version. Using the original version for all three study years is represented by DPAX111 and using the second version in the last two study years is represented by DPAX022.

Justification for the DPAX122 model as it pertained to depression stemmed from four factors (11):

1. DPAX-1 gave steady prevalence comparing 1952 and 1970; DPAX-2 gave steady prevalence comparing 1970 and 1992;
2. The same results about the prevalence of depression and its relationship to gender and age were produced for 1970 irrespective of which DPAX version was used;
3. The case-by-case agreement about depression between the two versions was good in 1970 and not as good in 1992;
4. DPAX-1 showed a decline in the prevalence of depression in 1992 probably due to the changing vernacular about depressed mood.

These factors suggested that 1970 could be thought of as a transitional year in which either version could be used. In other words, essentially the same results were produced irrespective of which version of DPAX was used for that year. This implied that it was after 1970 that 'poor spirits' began to become obsolete. We chose DPAX-2 for 1970 because it gave more options for conveying information about the 'essential features' and 'impairment' while adhering to the same step-by-step algorithm for diagnosis. In this paper, however, we evaluate the DPAX122 model as it applies to anxiety by the same four points used for depression as well as by qualitative interpretation of the congruence with clinical criteria.

The interview schedule asks subjects to report whether they experienced a given symptom 'often', 'sometimes', or 'never'. For the components in DPAX-1 where only one question is available, a response of 'often' or 'sometimes' indicated that the requirement of that particular step had been met. Where more than one question was available, the requirements of that step were met by a frequency-weighted threshold score with 'often' being valued as 3, 'sometimes' as 2, and 'never' as 1.

Both versions allow the calculation of current and lifetime prevalence. In DPAX-1, time of occurrence and duration are based on the time the subject was impaired, and in DPAX-2 they were based on the time when the 'essential symptoms' were experienced. While the minimum duration for both depression and anxiety is 1 month, the actual durations tended to be much longer. Across the study years, approximately half of all episodes lasted longer than 5 years, with those by DPAX-2 tending to be somewhat longer than those by DPAX-1. Pertinent to the sequential relationship between the two disorders is the fact that anxiety usually preceded depression.

Acute and chronic depressions are not differentiated in DPAX and thus the definition can be thought of as a combination of 'depressive

episodes' and 'dysthymia' as used in the International and American nomenclatures (64, 65). The definition of anxiety approximates that for Generalized Anxiety Disorder as it has traditionally been defined (Table 3). The associated symptoms of autonomic hyperactivity and motor tension upon which the definition concentrates are free-floating and are not required to appear in 'attacks', as in panic disorder. Our interview schedule also does not include questions about those anxiety disorders that involve phobic and obsessive–compulsive symptoms.

In our investigation of depression, we used the hierarchical principle of diagnosis and counted all cases of depression irrespective of whether anxiety was also present (11). In this report, we show prevalence in several ways. To convey the levels of comorbidity we use bar graphs showing the prevalence of depression alone, depression with anxiety, and anxiety alone. In tabular form we present prevalence rates for all anxiety, all depression, and all depression and anxiety disorders aggregated together.

Statistical methods

For the *methodologic* aim of assessing the DPAX122 model for anxiety, we used methods designed by Fitzmaurice et al. and Horton et al. for analyzing reports from multiple sources of information (71–74). These methods are equivalent to fitting simultaneously two logistic regression equations, one using reports of depression and anxiety from DPAX-1 and the other using similar reports from DPAX-2. The correlation between the two versions was accounted for in a unified regression model using an empirical variance estimator in a generalized estimating equation framework (75, 76). Implemented in SAS PROC GENMOD, these procedures allowed assessment of the two DPAX versions in the transition year of 1970. We also calculated the kappa (κ) statistic as a measure of agreement between the two versions for that year. In order to see the differences in results that might obtain from using the three possible approaches, we conducted prevalence analyses for Model DPAX111 and Model DPAX022 as well as the proposed Model DPAX122.

For the aims of analyzing time trends in *comorbidity*, *prevalence*, and *distribution*, standard logistic regression models were employed using SAS PROC LOGISTIC to fit models for main effects and pairwise interactions (77). For prevalence the main outcome variable was current anxiety (with or without depression). The level of co-occurrence was assessed by employing reports of co-existing depression as a predictor variable. Three other predictor variables were used: year of study as a three-level class variable (1952, 1970, and 1992), gender, and age (<45 vs. $45+$). Non-significant interactions ($P > 0.10$) were removed from the model.

To facilitate interpretability, some of the results such as those pertinent to gender are presented as odds ratios (OR) with 95% confidence intervals (CI). In keeping with the fact that an international task force has recommended that information about the comorbidity of two disorders be presented by OR (19), we calculated OR for this purpose as well. The findings in our study suggested, however, that the relationships between the two disorders could be more clearly seen by using predicted probabilities.

Prevalence rates were standardized by the direct method utilizing, as the external standard, official census information about all adults residing in the county, combining the 3 years of study (78).

A separate analysis was carried out that focused only on prevalence among women. Observation of our findings about depression led us to hypothesize a 'birth cohort effect' among women but we did not test it at the time. Most of the recent work on 'birth cohort effects' has used survival analysis of retrospective incidence from a sample interviewed at one point in time and has shown the results by means of graphs of cumulative lifetime rates for different cohorts (7–10). Because our data derive from three successive samples, we used a logistic regression model to test for an interaction between age and year of study to see if there was an significant increase among younger women in 1992 where both depression and anxiety were concerned.

Results

In regard to the *methodologic* aim of assessing DPAX122 for anxiety using the four points developed for depression, the results were as follows:

1. Holding DPAX version constant (DPAX111 and DPAX022), the analysis indicated that the prevalence of anxiety did not vary significantly by year of study in either model. The association of depression with anxiety was highly significant in each (with 1 d.f., chi-square values ranging from 173.2 to 288.5; and the level of significance was $P < 0.0001$ in each). The association between being female and the prevalence of anxiety was also significant in each (with 1 d.f., chi-square values ranging from 22.1 to 28.1; and the level of significance was $P < 0.0001$ in each). The association

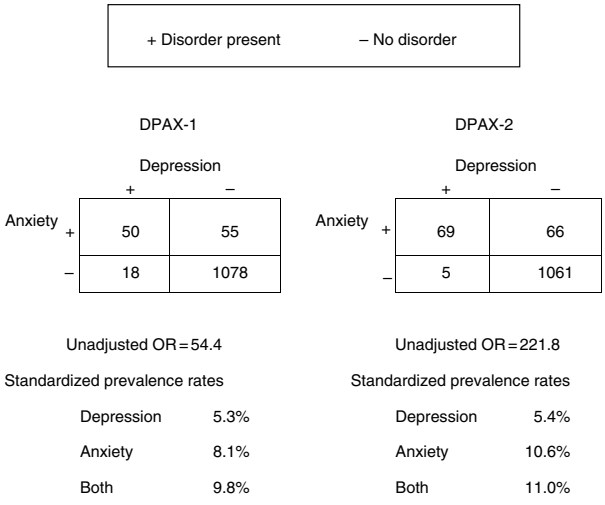

Fig. 1. Association of depression and anxiety comparing two versions of DPAX among 1201 subjects interviewed in 1970. DPAX-1 and DPAX-2: original and revised versions of a data-gathering procedure and diagnostic algorithm named Depression and Anxiety (DPAX). There were no significant differences between the two versions in regard to associations with age and gender but DPAX-2 showed a significantly higher level of comorbidity between depression and anxiety than did DPAX-1 ($\chi^2 = 6.0$, d.f. $= 1$, $P = 0.014$).

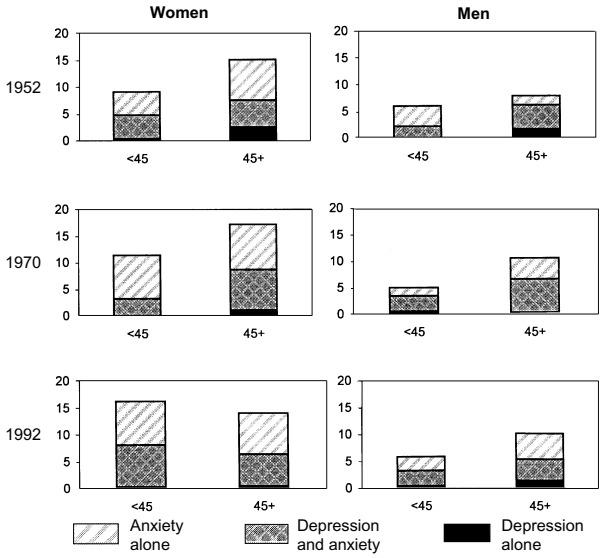

Fig. 2. Standardized cumulative prevalence rates (per 100) for anxiety alone, depression and anxiety, and depression alone by gender and age for three study years.

between anxiety prevalence and age was not so clearly comparable.

2. In regard to the proposed transition year of 1970 (Fig. 1), the simultaneous logistic regression model indicated that there were no significant differences between the two versions in regard to age and gender. There was, however, one significant difference in that the comorbidity of depression and anxiety was stronger in DPAX-2 than in DPAX-1.
3. The case-by-case agreement about anxiety between the two versions was good in 1970 ($\kappa = 0.55$) and not as good in 1992 ($\kappa = 0.47$).
4. Although not significant, as it had been for depression, DPAX111 showed a decline in the prevalence of anxiety (8.4% in 1952, 8.1% in 1970 and 6.2% in 1992).

These results indicated that the DPAX122 model functioned in essentially the same way for anxiety that it had for depression. Therefore, we used the DPAX122 model for the remainder of the analysis.

In regard to the aim of investigating the *comorbidity* of the two syndromes, we found that the presence of depression significantly influenced the presence of anxiety but that the prevalence of anxiety alone was consistently greater than that for depression alone (Fig. 2). Taking into consideration each age and gender category in each year, the predicted probabilities of having anxiety for subjects who reported depression ranged from 0.54 to 0.98. Among those who did not report depression, the probabilities ranged from 0.03 to 0.10 (Table 4). There were two significant interactions

Table 4. Estimated probability of anxiety in three samples by age, gender, and depression status*

	Year of sample selection					
	For subjects reporting depression			For subjects not reporting depression		
	1952	1970	1992	1952	1970	1992
Women						
<45	0.91	0.98	0.97	0.05	0.07	0.08
45+	0.72	0.94	0.91	0.07	0.09	0.10
Men						
<45	0.82	0.97	0.94	0.03	0.03	0.04
45+	0.54	0.88	0.82	0.03	0.04	0.05

* The comorbidity of the two disorders was significantly greater among younger than older subjects ($\chi^2 = 7.0$, d.f. $= 1$, $P = 0.008$). The association between the two disorders was also significantly stronger in 1970 than in 1952 ($\chi^2 = 7.4$, d.f. $= 2$, $P = 0.024$).

Table 5. Standardized prevalence rates per 100 (standard errors) of all anxiety disorders, all depression disorders, and their aggregation by gender and age in three study years*

	Women			Men			All
	<45	45+	Total	<45	45+	Total	Total
1952							
All anxiety	8.7 (1.7)	12.6 (2.1)	10.7 (1.4)	6.0 (1.7)	6.3 (1.6)	6.1 (1.2)	8.4 (0.9)
All depression	4.8 (1.3)	7.5 (1.7)	6.2 (1.1)	2.2 (1.1)	6.2 (1.5)	4.2 (0.9)	5.2 (0.7)
Aggregated disorders	9.1 (1.8)	15.0 (2.3)	12.2 (1.4)	6.0 (1.7)	7.9 (1.7)	7.0 (1.2)	9.6 (0.9)
1970							
All anxiety	11.3 (2.0)	16.1 (2.1)	13.8 (1.4)	4.4 (1.3)	10.2 (1.7)	7.3 (1.1)	10.6 (0.9)
All depression	3.2 (1.1)	8.6 (1.6)	6.0 (1.0)	3.3 (1.2)	6.5 (1.4)	4.9 (0.9)	5.4 (0.7)
Aggregated disorders	11.4 (2.0)	17.1 (2.1)	14.3 (1.5)	4.9 (1.4)	10.5 (1.7)	7.7 (1.1)	11.0 (1.3)
1992							
All anxiety	15.9 (2.1)	13.6 (1.7)	14.7 (1.3)	5.4 (1.4)	8.8 (1.5)	7.1 (1.1)	10.9 (0.9)
All depression	8.2 (1.1)	6.3 (1.2)	7.2 (1.0)	3.2 (1.1)	5.3 (1.2)	4.2 (0.8)	5.7 (0.6)
Aggregated disorders	16.3 (2.1)	13.9 (1.7)	15.1 (1.4)	5.8 (1.5)	10.1 (1.6)	8.0 (1.1)	11.5 (0.9)

* The presentation of time trends in this table makes use of the original version of the Depression and Anxiety (DPAX) algorithm for 1952 and the updated version for 1970 and 1992. The effect of study year on the prevalence of anxiety was not significant but women were significantly more likely to be diagnosed as anxious than men (OR = 2.2, 95% CI = 1.6–2.6). Among women, there was a significant interaction between age and study year indicating a higher rate in 1992 among younger women than in earlier years for depression ($\chi^2 = 6.7$, d.f. = 2, $P = 0.03$) and for the aggregated rate ($\chi^2 = 6.0$, d.f. = 2, $P = 0.048$) but not for anxiety. If the original version of DPAX had been used for all years, the prevalence of anxiety would have declined, although not significantly, from 8.4 to 8.1 and to 6.2%.

regarding the affiliation between depression and anxiety. One indicated that comorbidity was more pronounced among younger than older people. The other showed that the association of the two syndromes was stronger in 1970 than in 1952, a fact that relates to the transition from DPAX-1 to DPAX-2.

Regarding the aim of assessing time trends in the *prevalence* of anxiety, we found that the effect of study year was not significant indicating that, like depression, the overall rate of anxiety remained stable over time (Table 5, Fig. 2). It can be seen, however, that anxiety was more common than depression. The overall prevalence of anxiety ranged from 8.4 to 10.9% while that for depression ranged from 5.2 to 5.7%. Due to the high level of comorbidity, the aggregated rates representing all subjects who suffered from either depression and/or anxiety were quite similar to those for all anxiety in that they ranged from 9.6 to 11.5%. The prevalence of those who suffered from anxiety alone tended to be comparable with those who suffered from depression with or without anxiety.

In terms of the aim of assessing the *distribution* of these disorders according to gender and age, we found that women were consistently more than twice as likely to be diagnosed as having an anxiety disorder than were men while depression was not significantly different comparing men and women until the most recent sample (Table 5, Fig. 2). Where men were concerned, the rates of both anxiety and depression remained very similar over time both in overall terms and in terms of age distribution. In regard to the distribution of these disorders by age among women, we found that study year and age interacted significantly for depression as shown in a higher rate of depression among younger women in 1992 than in 1952 and 1970 ($P = 0.03$). While anxiety followed the same pattern to some extent, the interaction was not significant.

Discussion

The four components of this study can be summarized in the following way:

1. In regard to the *methodologic* aim of assessing a time trend strategy, analysis indicated that it functioned for anxiety about as well as it had for depression.
2. In regard to the *comorbidity* of depression and anxiety, the two disorders were found to be strongly and steadily associated with each other.
3. In regard to *prevalence*, the relationships between depression and anxiety did not change over time. Depression did not replace anxiety, and both disorders exhibited stable prevalence with the rate of anxiety being consistently twice as high as that of depression.
4. In regard to *distribution*, the prevalence of anxiety was significantly and consistently higher among women than among men while it was not until 1992 that depression was similar in this regard. The similarity was achieved through an increase of depression among younger women.

Methodological aim

In regard to this aim, it is worthwhile to review some of the difficulties involved in trying to assess the influence of the passage of time in regard to the goals of the study and to comment on the methodologic approaches we used to try to overcome them (28).

During the period when the study was conducted, different schools of psychiatric thought emerged, disappeared, or changed. The diagnostic practices of the psychiatrists trained in these schools also varied, with some being more oriented toward psychological theory and some more toward observation of symptoms. The discovery and use of psychotropic drugs also affected diagnostic practices, and responses to such medications had further impact on how the disorders were conceptualized. Public awareness of psychiatric disorders increased, and the words employed in talking about them also changed. Then in the last two decades, international effort was brought to bear on developing agreed-upon criteria for making psychiatric diagnosis more reliable. The availability of such standard definitions generated considerable research, some of which cast doubt on the criteria, which in turn led to further reconceptualization. It can be asked, therefore, whether it is possible at all to assess the influence of the passage of time on the comorbidity, prevalence and distribution of anxiety and depression.

One of our main strategies for overcoming the variability involved in the different schools of psychiatric thought was to focus on descriptive definitions based on the symptomatic phenomena of anxiety and depression rather than theoretical formulations about them. We found that psychiatrists could reach consensus under such circumstances although they did not agree about the etiological factors that may be involved. Whether subscribing to the psychodynamic orientation in which unconscious anxiety was a fundamental concept or to a biogenetic orientation, for example, psychiatrists were able to agree that anxiety was manifest when certain types of symptoms were reported.

The fact remains, however, that the value of such a phenomenal approach to diagnosis depends on whether the phenomena addressed by the questions asked refer to anxiety and depression as generally recognized. Looking to texts of psychiatry as well as the efforts to systematize psychiatric classification, it seems to us that for the greater part of the twentieth century, our descriptive definitions met such a requirement (79–82).

Our definition of depression as a disorder of lowered mood associated with disturbances of sleep, appetite and energy remains fairly well coordinated with contemporary standards although it does not include the recently added atypical disturbances of hypersomnia and increased appetite and it does not explicitly include self-deprecation and thoughts of suicide (35).

Where anxiety is concerned, our definition resembles that for generalized anxiety disorder in the current international classification and in DSM-III as a disorder in which fearful apprehension is accompanied by sweating, palpitations, trembling and other indicators of autonomic hyperactivity and motor tension (35, 64). This definition is, however, quite different from that in DSM-IV (65). Recent research in the US indicated that generalized anxiety in these traditional terms was associated with each of the specific forms of anxiety – panic, obsessive–compulsive, and phobic disorders (83, 84). This suggested that if generalized anxiety were to be sustained as an independent disorder, a different definition might be needed. The new definition given in DSM-IV requires chronic and excessive worry focused on multiple life situations and involves such associated symptoms as tension, restlessness, and sleep disturbance but it specifically excludes autonomic hyperactivity. If this new definition gains international acceptance, a future survey using our questions (which tend to concentrate on autonomic hyperactivity) would not identify what by then might be called generalized anxiety. As our goal was to characterize the 40 years from 1952 to 1992, however, we suggest that what we mean by depression and anxiety was fairly well accepted during that period.

Another approach we took to the problem of variability in psychiatric assessment was to develop computerized algorithms for diagnosis. Such computer-implemented procedures are useful because they are consistent and do not drift away from the standard definition over time. They also impose on the analysis a level of explicitness that is sometimes difficult to achieve when diagnosis is carried out only by human clinical assessment. Thus it is possible to show the questions asked and the algorithms by which the responses were analyzed so that they can be evaluated and interpreted.

Another feature of our effort to present as accurate an assessment of time trends as possible was to study samples of adults selected at different points in history and to focus on the subjects' responses about their current clinical state at those times. Such an approach avoids the possible distortion involved in using recalled information

as when time trends are based solely on studying a sample selected at one point in time and asking subjects to recollect their experiences with anxiety and depression.

With the passage of time, however, changes occur in the vernacular language employed by the population at large for describing their feelings. To address this issue, we developed a strategy that involved increasing the number of questions in the interview schedule so as to use more and updated ways of asking about the central symptoms of anxiety and depression. It is well known, however, that changes in an interview schedule can produce changes in the rates to be reported (85).

In light of this, it would have been desirable to have constructed an interview schedule at the beginning of the study that contained several questions about dysphoria and apprehension and to have been able to cast them in idioms that would resist vernacular change. Such forecasting of the future is, however, exceedingly difficult. As an alternative, our approach was to add, at the midpoint of the study, a few extra questions with the expectation that the collection would include words that would be understandable to people of many different walks of life at different points in history and to use two versions of the DPAX algorithms seriatim over the course of study.

In evaluating the strategy of using DPAX-1 for 1952 and DPAX-2 for 1970 and 1992, our assessment indicated that this approach functioned in essentially the same way for both depression and anxiety thus making it possible to conduct the analysis in a unified way. The features that most recommended the strategy were that DPAX-1 gave stable prevalence of both disorders for 1952 and 1970 and DPAX-2 gave stable prevalence of both disorders for 1970 and 1992, and that for the transition year of 1970 the case-by-case agreement between the two versions was good and the same results about prevalence and its relationships to age and gender were produced irrespective of which version was employed.

Nevertheless, two aspects of the application for anxiety differed from depression. One was that DPAX-2 showed a higher level of comorbidity than DPAX-1 although such comorbidity was exceedingly strong in both versions. The other was that secular drift in the words for anxiety was not as clear-cut as for depression. There was no dramatic decline in the prevalence of any of the anxiety symptoms cataloged in DPAX-1 as there had been for 'poor spirits'.

We wondered if using the idioms of 'nervousness' and 'nervous breakdown' might have become outdated, but our evidence suggests that they continued to be useful ways for people to convey a state of feeling that mainly seems to reflect anxiety and, when allied with 'breakdown', to convey apprehension about ability to maintain emotional equilibrium. The proportions of the samples that gave positive responses about 'nervous breakdown' remained steady across the three samples, and positive responses about 'nervousness' increased.

Most of the recent manuals and interview schedules use 'nervousness' either in regard to anxiety specifically or to emotional problems generally. This seems to support the view that the idea of 'nerves' continues to have value in communicating about psychiatric phenomena although the general trend of the times is toward using more specific diagnostic terms. One can forecast, however, that there may come a time when 'nervousness' will have disappeared from the vernacular in the same way that melancholy, poor spirits, and 'feeling spleeny' seem to have become old-fashioned. The CIDI, for example, uses the words 'anxiety' and 'worry' in places where other schedules have traditionally used 'nervousness'. This suggests that like the word 'depression', the word 'anxiety' is now sufficiently well understood that it can be used in general population surveys. It is possible, however, that without also using 'nervousness', contemporary research could be biassed because of employing words that are more familiar to educated or younger people than those who are older or uneducated.

The response that mainly influenced the decline of prevalence regarding anxiety according to DPAX-1 was 'going easy', which like 'poor spirits', had the disadvantage of being the only question available for one of the algorithm steps. In order to avoid having only one question for these important steps, DPAX-2 involves a larger number of questions than DPAX-1. Theoretically, this could lead to a larger number of positive responses and therefore to the possibility of artificially higher rates. The factors that appear to have restrained such a tendency are that: (i) the overlap of questions between the two versions is extensive as only one question in DPAX-1 is not in DPAX-2 ('going easy on work'); (ii) the added questions concerned themes that existed in the original; (iii) most of the questions added for impairment were asked only of those subjects who gave positive evidence about the 'essential features'; (iv) the analysis of both the original and updated versions followed the same algorithmic steps; and (v) the algorithms reproduce decisions made by psychiatrists who probably equated responses on similar themes and therefore offered

a way of balancing the number and meaning of positive responses.

Because of concern about the validity of our methods, we conducted three investigations in which the survey results were assessed by comparing them to decisions provided by psychiatrists. While such an approach to validity is a common one for epidemiologic investigations, it is generally recognized that a perfect 'gold standard' does not exist and that while the reliability of psychiatric diagnoses has been improved by application of the procedures spelled out in the recent classification systems, validity is still problematic.

Our efforts to assess validity reflect methodological changes in the field of psychiatric epidemiology as a whole. This is illustrated in the fact that our methods began by having the psychiatrist use an unstructured approach leading to a decision as to whether the subject was a 'psychiatric case' or not and ended by having clinicians use the SCID and provide specific diagnoses. The best results as indicated by sensitivity and specificity pertained in the study carried out in the 1960s when the survey data included both self-report and reports by general physicians and when the assessment did not focus on specific diagnoses but rather on whether the survey materials and the psychiatrists agreed that the subject had a psychiatric disorder of some kind. The procedures of the human psychiatric evaluation were those reproduced by DPAX-1. Thus the findings give credence to the idea that DPAX-1 seemed to functioning adequately at that point in time.

The evolution of our methods for assessing validity involved an increasingly difficult task. Agreement about a specific diagnosis constitutes a more stringent test than does recognition that a psychiatric problem exists. It seems likely that this increased difficulty relates to the fact that the validity evidence from the recent study is less good than that from the earlier studies.

However, the sensitivity and specificity results based on our use of SCID are quite comparable with those produced in a number of other recent studies of validity that have focused on specific diagnoses (86–88). Our best results pertained for the combination of anxiety and depression, followed by anxiety, and the worst were for depression. Where depression was concerned, the lowest sensitivity pertained for DPAX-1, the next higher was for DPAX-2, and the highest was for the DIS. We interpret this as offering further evidence that DPAX-1, with its dependence upon 'poor spirits' was becoming outdated and thus supporting our strategy of transitioning to DPAX-2. The better agreement for the DIS may relate to the fact that its language and that of SCID are very similar, thus adding weight to the view that the language by which depression is identified makes a difference in terms of such agreement. Where depression is concerned, it is important to emphasize that each of the three survey procedures (DPAX-1, DPAX-2, and DIS) gave high specificity values while at the same time giving low sensitivity. This means that it was very rare for the clinicians using SCID to judge that a person identified as depressed by any one of the survey methods was *not* depressed. The clinicians accepted almost everyone identified as depressed by the survey methods as indeed being depressed but at the same time they identified many subjects as depressed who had not been so identified by the survey methods. This means either that the survey methods underestimated the prevalence of depression or the SCID methods overestimated such prevalence. Extrapolating from the validity results, the magnitude of the difference can be estimated as ranging from 40% lifetime prevalence based on the clinicians' assessments using SCID to 8 or 11% lifetime prevalence for DPAX-2 and the DIS, respectively. This result was very surprising because it has frequently been suggested that psychiatric epidemiologic studies overestimate prevalence by counting transient disturbances of mood and that such inaccuracy would be demonstrated if clinicians provided the assessments (85).

About the same time that we were carrying out the SCID validity investigation in Stirling County, a similar effort was conducted as part of the Baltimore ECA Follow-up Study (89). The focus was on depression as measured by the DIS and the clinical standard was the World Health Organization's Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (90). As in our study, the clinical methods identified a much higher rate of depression than did the epidemiologic methods.

Such findings raise questions about the validity of both clinical and epidemiologic methods, perhaps especially for depression. For example, does the amount of attention given to depression in recent years have something to do with why the clinicians saw so much of it? Given such questions, it would be helpful to study predictive validity so as to address the question of whether one or both types of methods tended to identify transient aberrations of mood rather than the kinds of chronic disorders that have often been found characteristic of clinic patients (91).

Regarding long-term course and outcome, two types of follow-up information from our study are pertinent. In the validity investigation conducted in the middle phase of our study, psychiatrists continued to visit the subjects in their homes each year

for 5 years. Using the psychiatrists' assessments, it appeared that over 50% of the cases identified by the survey methods remained ill during that period (54). In following cases identified by DPAX-1 over a 16-year period, the re-interview data among survivors indicated that over 75% of those depressed at baseline had a poor outcome in terms of remaining chronically or recurrently depressed or having the often comorbid anxiety disorder become predominant (42). We also found that depression was associated with premature mortality, especially among men (43). Such evidence has encouraged us to believe that the phenomena identified by our survey methods are not ephemeral disturbances but rather serious disorders similar to those seen in psychiatric treatment facilities. Until similar follow-up studies can be carried out based on the clinician-recognized cases of depression, it will be difficult to resolve the questions raised by the SCAN and SCID studies in the Baltimore-ECA and Stirling investigations.

Aim regarding comorbidity

We found that depression and anxiety were strongly and consistently affiliated with each other. The pattern of association, however, was lop-sided in that it was rare to find depression alone but common to find anxiety alone. This result differs from the more balanced relationship reported in studies that have used the DIS or CIDI (15, 19). The difference probably follows from the way in which the questions are asked.

In the DIS and CIDI, each disorder is taken up in a separate unit of the interview. Such units focus on discrete 'episodes', the minimum duration of which is built into the questions, as in, 'Over your lifetime has there ever been a 2-week period when you felt…', and for which care is taken to insure that all features of the disorder were in existence during the 'episode'. Such emphasis on specific periods of time reduces the likelihood that an episode of depression will overlap with an episode of anxiety.

In our interview, the questions about depression and anxiety are intermixed and dating takes place at the end when the subject is asked to give whatever durations are personally relevant. Using such an approach, the durations in our study are also longer than those customarily reported in DIS and CIDI studies (92, 93). This feature also contributes to our finding that depression is almost always accompanied by anxiety as long durations are more likely to foster the co-existence of two disorders at any one point in time than are discrete episodes defined in terms of a minimum duration.

In addition to the fact that rather long durations have been given in an initial interview with a subject in our study, the re-interviews conducted in follow-up investigations have also indicated considerable chronicity (41, 42). It would be a mistake, however, to interpret the durations in our study as referring to constant states of florid illness. Our methods do not permit refined depiction of the fluctuations of illness that in all probability pertain.

Notwithstanding the fact that the definition of generalized anxiety may change in the future, we interpret our findings about the association between depression and anxiety as supporting the view that anxiety as we measured it is a distinctive disorder rather than the view that it is an incomplete disorder whose sole significance is as a premonitory or residual phenomenon. In part, this interpretation stems from the fact that, in terms of current prevalence, anxiety alone was as common as depression with or without anxiety. Further, those who were diagnosed as having anxiety alone met the same criteria for syndrome completeness, duration, and impaired functioning as those diagnosed as depressed. The distinctiveness of the affective states and physiological concomitants that express the two disorders should also not be ignored. The fearfulness, sweating, trembling, palpitations that subjects report when diagnosed as having an anxiety disorder are different from the sadness, loss of energy, sleep, and appetite that they report when diagnosed as depressed. If anxiety is a distinct disorder, its correlates should differ from those of depression. In this regard, we found that anxiety was more consistently and strongly a disorder of women than was true for depression.

From the perspectives of both prevalence and incidence, we found that anxiety usually preceded depression, which is a sequence that has been reported in numbers of other studies (94–98). Further, the follow-up of the depressed men and women identified in our first sample indicated that men tended to remain chronically or recurrently depressed while women were more likely to lose depressive features and have the previously comorbid anxiety become dominant (99). Although such evidence might suggest that general anxiety was both the prodrome and postdrome of depression, even in such 'before' and 'after' occurrences, anxiety met its own diagnostic criteria just as much as depression met its criteria.

It has been suggested that a phenomena called 'double depression' is a useful way to describe the

course of chronic depression where dysthymia is the underlying disorder with depressive episodes superimposed on it (100). An image that characterizes the sequential co-existence of anxiety and depression in our study is one of disorders piling up on top of each other. A typical pattern is for anxiety to appear first and to persist. Then chronic depression may come into existence and during its course may exhibit several fluctuations in level of severity. If the depression resolves, the anxiety often continues as an impairing condition. It seems to us that such a pattern of affiliations warrants the view that the relationship is one of the comorbidity of distinctive disorders.

Aim regarding prevalence

Our study did not indicate that depression replaced anxiety nor that there was an overall increase pertinent to either or both. On the face of it, these findings may seem incompatible with the evidence presented by those researchers who have used the concepts of 'age of melancholy' and 'age of anxiety'. We suggest, however, that incongruity is not necessarily the case.

Klerman's use of 'age of melancholy' came partly from patients who sought help through treatment facilities (5). It is well established that there are marked differences between 'true' and 'treated' prevalence. It is quite conceivable that despite the steady prevalence of depression in a general population as reported here, increasingly large numbers of people have sought treatment for depression, especially as such treatment has become more available and as more practitioners have become capable of recognizing depression.

In addition, Klerman pointed out that a change of attention on the part of clinicians seemed to occur. Anxiety appeared to be a more central focus of clinical concerns in the period following World War II, in part due to the role played by unconscious anxiety in psychoanalytic theory. Depression was more central after the antidepressant medications came into common usage around 1970. Further, this shift in the focus of attention seemed to be reflected in the population at large as well as among clinicians. In our study, numerous indicators pointed to increased awareness and acknowledgement of depression. For example, our interview schedule begins with open-ended questions about health and, during the 40-year period, numbers of subjects have described psychiatric illnesses in this section. In 1952, when 4% volunteered such conditions, only one person used the word 'depression' while all the others used some variant of 'nervous trouble' (58). By 1992 when 9% described such conditions, the language used was more specific and clinically sophisticated. By that time 'depression' was the second most commonly used word although 'nervous trouble' continued in first place.

The evidence used by Twenge regarding an increase in anxiety was based on symptom levels (4). Focusing on symptoms *per se* in our study, we pointed out that some of the pivotal questions involved idioms that seemed to lose currency over time but positive responses to several of the other questions increased (58). On the whole, increase at the level of individual symptoms was more prominent than decrease, especially for symptoms of lowered mood and nervous apprehension. This seems to fit well with mounting evidence that recent generations are more 'psychologically minded' than their predecessors (101, 102). Changes in symptom levels are not, however, necessarily incompatible with steady prevalence of rates of disorder when, as in the DPAX algorithms, all the parameters of a diagnosis are taken into account, including clinical significance as indicated by the duration and intensity of symptoms and their interference with ordinary life.

Similarly, a contradiction does not necessarily exist between a stable current prevalence rate and an increase in incidence. This is because prevalence may be especially influenced by long-term disorders and incidence by acute or highly episodic disorders. While there are a growing number of long-term incidence studies (103, 104), to the best of our knowledge the Lundby and Stirling Studies are the only ones that have thus far been able to assess the effect of the passage of time by comparing an early follow-up study with a later one (3, 14). Despite the fact that the sites of the Lundby and Stirling studies are quite similar, Lundby reported an increase in depression incidence while we reported steady rates for both the incidence and prevalence of depression. In regard to both types of rates, we suggest that the difference may be due to the methods used. The Lundby approach of having psychiatrists interview subjects may be particularly sensitive to short-lived episodes while our methods are more sensitive to chronic disorders.

Aim regarding distribution according to gender and age

In initiating the work on the prevalence of anxiety, one of our interests was to learn whether there were changes in the distribution of anxiety according to age and gender that might resemble what we interpreted as a 'birth cohort effect' among women regarding depression. Because we had not earlier tested the possibility of an interaction between the

passage of time and age among women, we included such analysis in this study. We found that the interaction was significant for depression and that while anxiety showed something of the same trend it was not significant.

One of the strong findings about the prevalence of anxiety was its being consistently higher among women than men. The increase in depression among younger women meant that by the end of the study the difference between men and women regarding depression was significant where it had not been so earlier. In other words, we found that depression came to resemble anxiety in this regard rather than anxiety resembling depression in a 'birth cohort effect'. It bears emphasis that the only change we found in the relationships between depression and anxiety was this one about distribution in which depression began to look more like anxiety.

We suggest that there is fairly direct congruence between our findings about the recent increase of depression and similar evidence from other studies. The different perspectives offered through retrospective incidence, prospective incidence, and ours through current prevalence seem to converge in pointing to recent increases of depression, at least among women. The studies that point in this direction have been carried out in different parts of the world and have used different methods. In so far as the validity of a finding is partly gauged by its being derived from different methods, the accumulating evidence strengthens the likelihood that these disorders are, indeed, gradually increasing.

Incidence studies are especially useful for giving evidence about who becomes afflicted with a disorder and when this occurs. While incidence studies thus give information about the onset of disorder they do not tell us anything about the duration of illness. Prevalence studies, however, convey information about the persistence of the disorder. Our findings about current prevalence suggest that, at least for women, the factors at work in producing the increase in depression seem to have contributed not only to the *onset* but also to the *duration* of depression.

Depression and anxiety, like other psychiatric disorders, are etiologically complex with contributions from both genetics and environment. It seems likely that the cases counted in a prevalence rate consist of some that have more input from genetics and others with more input from the environment. In view of this complexity, rapid and wholesale changes in rates are not likely to occur. We suggest that overall steadiness of rates is in line with etiological complexity. A gradual change, and perhaps especially one limited to a certain segment of the population, also has greater credibility than one that becomes widespread suddenly.

A change in the distribution of a disorder is caused by something. In the present case, knowledge of that 'something' is lacking. Because the increase in the Stirling Study impinged mainly on women who were born after the World War II, several hypotheses about alterations in the social environment may be suggested.

It was during the World War II that women began to enter the labor force in large numbers. After that war, family planning became more practical, and women began to bear smaller numbers of children. New expectations were held about how women should contribute to family and society as wives, mothers, and workers. The structures of society also began to change. Social class as traditionally measured began to have a less determining influence on attitudes and lifestyles (105, 106). The strength and support of religion and community gave way to an emphasis on the importance of the individual (107, 108). These and numerous other social changes may contribute to altering the epidemiologic profile of a population, some of which we will be able to investigate through our database.

The overall conclusion we draw from our study is that with one exception the relationships between anxiety and depression remained remarkably similar over time. A change in a relationship, however, is noteworthy because it may open a window of opportunity for exploring its correlates and possibly its causes. We interpret, therefore, that one of the most important findings of our investigation is that the distribution of depression changed so that it came to resemble the distribution of anxiety as disorders to which women are especially vulnerable, and that this vulnerability as it pertained to depression appeared mainly among those born after the World War II. Such evidence, especially when laid along side other studies that show recent increases in depression, raises questions about the possible relevance of social and cultural changes in the environment. It seems clear that more knowledge is needed about that 'something' which may relate to the increases that appear to be occurring in several parts of the world.

Acknowledgement

This research was supported by grant MH39576 from the US National Institute of Mental Health of the National Institutes of Health (Dr Murphy). The authors gratefully acknowledge the assistance of Patricia Merritt, Barbara Burns, and Ellen Krystofik for field management and data processing and of

Barbara Burns additionally for literature review and manuscript preparation.

References

1. Klerman GL. Affective disorders. In: Nicholi AM, ed. The Harvard guide to modern psychiatry. Cambridge, MA: Belknap Press of Harvard University Press, **1978**:253–281.
2. Klerman GL. The current age of youthful melancholia: evidence for increase in depression among adolescents and young adults. Br J Psychiatry 1988;**152**:4–14.
3. Hagnell O, Lanke J, Rorsman B, Öjesjö L. Are we entering an age of melancholy? Depressive illness in a prospective epidemiological study over 25 years: the Lundby Study, Sweden. Psychol Med 1982;**12**:279–289.
4. Twenge JM. The age of anxiety? Birth cohort changes in anxiety and neuroticism, 1952–1993. J Pers Soc Psychol 2000;**79**:1007–1021.
5. Klerman GL. Age and clinical depression: today's youth in the twenty-first century. J Gerontol 1976;**27**:318–323.
6. Robins LN, Regier DA. Psychiatric disorders in America: the Epidemiologic Catchment Area Study. New York: The Free Press, 1991.
7. Klerman GL, Weissman MM. Increasing rates of depression. JAMA 1989;**261**:2229–2235.
8. 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**:333–343.
9. Burke KC, Burke JD, Rae DS, Regier DA. Comparing age at onset of major depression and other psychiatric disorders by birth cohorts in five US community populations. Arch Gen Psychiatry 1991;**48**:789–795.
10. Cross-National Collaborative Group. The changing rate of major depression: cross-national comparisons. JAMA 1992;**268**:3098–3105.
11. Murphy JM, Monson RR, Laird NM, Sobol AM, Leighton AH. A forty-year perspective on the prevalence of depression from the Stirling County Study. Arch Gen Psychiatry 2000;**57**:209–215.
12. Kessler RC, Mcgonagle KA, Swartz M, Blazer DG, Nelson CB. Sex and depression in the National Comorbidity Survey, I: lifetime prevalence, chronicity and recurrence. J Affect Disord 1993;**29**:85–96.
13. Weissman MM, Bland R, Joyce PR, Newman S, Wells JE, Wittchen HU. Sex differences in rates of depression: cross-national perspectives. J Affect Disord 1993;**29**:77–84.
14. 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**:505–514.
15. Boyd JH, Burke JD, Gruenberg E et al. Exclusion criteria of DSM-III; a study of co-occurrence of hierarchy-free syndromes. Arch Gen Psychiatry 1984;**41**:983–989.
16. Maser JD, Cloninger CR. Comorbidity of mood and anxiety disorders. Washington, DC: American Psychiatric Press Inc., 1990.
17. Murphy JM. Diagnostic comorbidity and symptom co-occurrence: the Stirling County Study. In: Maser JD, Cloninger CR, eds. Comorbidity of mood and anxiety disorders. Washington, DC: American Psychiatric Press Inc., **1990**:153–176.
18. Kessler RC, McGonagle KA, Zhao S et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. Arch Gen Psychiatry 1994;**51**:8–19.
19. Merikangas KR, Angst J, Eaton W et al. Comorbidity and boundaries of affective disorders with anxiety disorders and substance misuse: results of an international task force. In: Wittchen HU, ed. Comorbidity of mood disorders. Br J Psychiatry 1996;**168(suppl. 30)**:58–67.
20. Sartorius N, Üstün B, Lecrubier Y, Wittchen HU. Depression comorbid with anxiety: results from the WHO study on psychological disorders in primary health care. In: Wittchen HU, ed. Comorbidity of mood disorders. Br J Psychiatry 1996;**168(suppl. 30)**:38–43.
21. Kessler RC, Dupont RL, Berglund P, Wittchen HU. Impairment in pure and comorbid generalized anxiety disorder and major depression in 12 months in two national surveys. Am J Psychiatry 1999;**156**:1915–1923.
22. Wittchen HU, Hoyer J. Generalized anxiety disorder: nature and course. J Clin Psychiatry 2001;**62**:15–19.
23. Hunt C, Issakidis C, Andrews G. DSM-IV generalized anxiety disorder in the Australian National Survey of Mental Health and Well-Being. Psychol Med 2002;**32**: 649–659.
24. Angst J, Dobler-Mikola A. The Zurich Study: VI. A continuum from depression to anxiety disorders. Eur Arch Psychiatry Neurol Sci 1985;**235**:179–186.
25. Eaton WW, Ritter C. Distinguishing anxiety and depression with field survey data. Psychol Med 1988;**18**: 155–166.
26. Krueger RF. The structure of common mental disorders. Arch Gen Psychiatry 1999;**56**:921–926.
27. Breier A, Charney DS, Heninger GR. The diagnostic validity of anxiety disorders and their relationship to depressive illness. Am J Psychiatry 1985;**142**:787–797.
28. Isohanni MK. Editorial: Administrative aspects in longitudinal studies: how to navigate on a stormy and dangerous ocean? Acta Psychiatr Scand 2001;**104**:1–3.
29. Leighton AH: My name is legion: the Stirling County Study of psychiatric disorder and sociocultural environment. New York: Basic Books, 1959.
30. Hughes CC, Tremblay MA, Rapoport RN, Leighton AH: People of cove and woodlot: the Stirling County Study. New York: Basic Books, 1960.
31. Leighton DC, Harding JS, Macklin DB, Macmillan AM, Leighton AH: The character of danger: the Stirling County Study. New York: Basic Books, 1963.
32. Murphy JM, Sobol AM, Neff RK, Olivier DC, Leighton AH. Stability of prevalence: depression and anxiety disorders. Arch Gen Psychiatry 1984;**41**:990–997.
33. Murphy JM, Olivier DC, Monson RR, Sobol AM, Federman EB, Leighton AH. Depression and anxiety in relation to social status: a prospective epidemiologic study. Arch Gen Psychiatry 1991;**48**:223–229.
34. Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics and validity. Arch Gen Psychiatry 1981;**38**:381–389.
35. American Psychiatric Association. Diagnostic and statistical manual, 3rd edn. Washington, DC: American Psychiatric Association, 1980.
36. Weissman MM, Bruce ML, Leaf PJ, Florio LP, Holzer C. Affective disorders. In: Robins LN, Regier DA, eds. Psychiatric disorders in America: the Epidemiological Catchment Area Study. New York: Free Press, 1991:53–80.
37. Regier DA, Boyd JH, Burke JD et al. One-month prevalence of mental disorders in the United States based on five epidemiologic catchment area sites. Arch Gen Psychiatry 1988;**45**:977–986.

38. Bland RC, Orn H, Newman SC. Lifetime prevalence of psychiatric disorders in Edmonton. Acta Psychiatr Scand 1988;**77(suppl. 338)**:24–32.
39. Bland RC, Newman SC, Orn H. Period prevalence of psychiatric disorders in Edmonton. Acta Psychiatr Scand 1988;**77(suppl. 338)**:33–42.
40. Murphy JM, Olivier DC, Monson RR, Sobol AM, Leighton AH. Incidence of depression and anxiety: the Stirling County Study. Am J Public Health 1988;**78**:534–540.
41. Murphy JM, Sobol AM, Olivier DC, Monson RR, Leighton AH, Pratt LA. Prodromes of depression and anxiety: the Stirling County Study. Br J Psychiatry 1989;**155**:490–495.
42. Murphy JM, Olivier DC, Sobol AM, Monson RR, Leighton AH. Diagnosis and outcome: depression and anxiety in a general population. Psychol Med 1986;**16**:117–126.
43. Murphy JM, Monson RR, Olivier DC, Sobol AM, Leighton AH. Affective disorders and mortality: a general population study. Arch Gen Psychiatry 1987;**44**:473–480.
44. Kessler RC, Berglund P, Demler O et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA 2003; **289**:3095–3105.
45. Murphy JM, Monson RR, Olivier DC, Sobol AM, Pratt LA, Leighton AH. Mortality risk and psychiatric disorders: results of a general physician survey. Soc Psychiatry Psychiatr Epidemiol 1989;**24**:134–142.
46. Essen-Möller E. Individual traits and morbidity in a Swedish rural population. Acta Psychiatr Neurol Scand 1956;**(suppl. 100)**:1–160.
47. Hagnell O. A prospective study of the incidence of mental disorder. Stockholm: Svenska Bokförlaget, 1966.
48. Hagnell O, Essen-Möller E, Lanke J, ÖJesjö L, Rorsman B. The incidence of mental illness over a quarter of a century: the Lundby longitudinal study of mental illnesses in a total population based on 42,000 observation years. Stockholm: Almqvist & Wiksell, 1990.
49. Star SA. Psychoneurotic symptoms in the army. In: Stouffer SA, Lumsdaine AA, Lumsdaine MH et al., eds. The American soldier: combat and its aftermath, Vol. II. Princeton, NJ: Princeton University Press, 1949:411–455.
50. Star SA. The screening of psychoneurotics in the army: technical development of tests. In: Stouffer SA, Guttman L, Suchman EA, Lazarsfeld PF, Star SA, Clausen JA. The American soldier: measurement and prediction, Vol. IV. Princeton, NJ: Princeton University Press, 1950:486–547.
51. Eysenck HJ. Dimensions of personality. London: Routledge & Kegan Paul, 1947.
52. Macmillan AM. The Health Opinion Survey: technique for estimating prevalence of psychoneurotic and other types of disorder in communities. Psychol Rep 1957;**3**: 325–339.
53. Leighton AH, Leighton DC, Danley RA. Validity in mental health surveys. Can Psychiatr Assoc J 1966;**11**:167–178.
54. Beiser M. A psychiatric follow-up of 'normal' adults. Am J Psychiatry 1971;**127**:1464–1472.
55. Beiser M. A study of personality assets in a rural community. Arch Gen Psychiatry 1971;**24**:244–254.
56. Murphy JM, Neff RK, Sobol AM, Rice JX, Olivier DC. Computer diagnosis of depression and anxiety: the Stirling County Study. Psychol Med 1985;**15**:99–112.
57. Murphy JM. Continuities in community-based psychiatric epidemiology. Arch Gen Psychiatry 1980;**37**:1215–1223.
58. Murphy JM, Monson RR, Laird NM, Sobol AM, Leighton AH. Identifying depression and anxiety in a forty-year epidemiologic investigation: the Stirling County Study. Int J Methods Psychiatr Res 1998;**7**:89–109.
59. Simpson JA, Weiner ESC (eds). The Oxford English Dictionary, 2nd edn. New York: Oxford University Press, 1989.
60. Benfari RC, Beiser M, Leighton AH, Murphy JM, Mertens CW. The manifestations of types of psychological states in an urban sample. J Clin Psychol 1974;**30**:471–483.
61. Jamison KR. Touched with fire: manic-depressive illness and the artistic temperament. New York: Simon & Schuster, 1993.
62. Winters EE (ed). The collected papers of Adolf Meyer, Vol III: medical teaching. Baltimore, MD: The Johns Hopkins Press, 1951.
63. Muncie W. Psychobiology and psychiatry: a textbook of normal and abnormal human behavior. St Louis, MO: C.V. Mosby Co., 1948.
64. World Health Organization. The ICD-10 classification of mental and behavioural disorders – clinical descriptions and diagnostic guidelines. Geneva: WHO, 1993.
65. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 4th edn. Washington, DC: American Psychiatric Association, 1994.
66. Robins LN, Wing J, Wittchen HU et al. The Composite International Diagnostic Interview. Arch Gen Psychiatry 1988;**45**:1069–1077.
67. Kessler RC, Wittchen HU, Abelson J et al. Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey (NCS). Int J Methods Psychiatr Res 1998;**7**:33–55.
68. Jenkins R, Bebbington P, Brugha T et al. The National Psychiatric Morbidity Surveys of Great Britain: strategy and methods. Psychol Med 1997;**27**:765–774.
69. Murphy JM, Monson RR, Laird NM, Sobol AM, Leighton AH. A comparison of diagnostic interviews for depression in the Stirling County Study: challenges for psychiatric epidemiology. Arch Gen Psychiatry 2000;**57**:230–236.
70. Spitzer RL, Williams JBW, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID), 1: history, rationale, and description. Arch Gen Psychiatry 1992;**49**:624–629.
71. Horton NJ, Laird NM, Zahner GEP. Use of multiple informant data as a predictor in psychiatric epidemiology. Int J Methods Psychiatr Res 1999;**8**:6–18.
72. Horton NJ, Laird NM, Murphy JM, Monson RR, Sobol AM, Leighton AH. Multiple informants: mortality associated with psychiatric disorders in the Stirling County Study. Am J Epidemiol 2001;**154**:649–656.
73. Fitzmaurice GM, Laird NM, Zahner GEP, Daskalakis C. Bivariate logistic regression analysis of childhood psychopathology ratings using multiple informants. Am J Epidemiol 1995;**142**:1194–1203.
74. Fitzmaurice GM, Laird NM, Zahner GEP. Multivariate logistic models for incomplete binary responses. J Am Stat Assoc 1996;**91**:99–108.
75. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;**73**:13–22.
76. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986;**42**:121–130.
77. Hosmer DW, Lemeshow S. Applied logistic regression, 2nd edn. New York: John Wiley & Sons Inc, 2000.
78. Anderson A, Auquier A, Hauck WW, Oakes D, Vandaele W, Weisberg HI. Statistical methods for comparative studies. New York: John Wiley & Sons Inc, 1980.
79. Jaspers K. General Psychopathology (translated from the German by Hoenig J, Hamilton MW). Manchester: Manchester University Press, 1963.

80. Ewalt JR, Farnsworth DL. Textbook of psychiatry. New York: McGraw-Hill Book Co., 1963.
81. Wing JK, Cooper JE, Sartorius N. Measurement and classification of psychiatric symptoms: an instruction manual for the PSE and Catego program. Cambridge, UK: Cambridge University Press, 1974.
82. Barlow DH: Anxiety and its disorders: the nature and treatment of anxiety and panic. New York: Guilford Press, 1988.
83. Barlow DH, Blanchard EB, Vermilyea JA, Vermilyea BB, DiNardo PA. Generalized anxiety and generalized anxiety disorder: description and reconceptualization. Am J Psychiatry 1986;**143**:40–44.
84. Barlow DH, Wincze J. DSM-IV and beyond: what is generalized anxiety disorder? Acta Psychiatr Scand 1998;**98(suppl. 393)**:23–29.
85. Regier DA, Kaelber CT, Rae DS et al. Limitations of diagnostic criteria and assessment instruments for mental disorders: implications for research and policy. Arch Gen Psychiatry 1998;**55**:109–115.
86. Anthony JC, Folstein M, Romanoski AJ et al. Comparison of the lay Diagnostic Interview Schedule and a standardized psychiatric diagnosis: experience in eastern Baltimore. Arch Gen Psychiatry 1985;**42**:667–675.
87. Helzer JE, Robins LN, Mcevoy LT et al. A comparison of clinical and Diagnostic Interview Schedule diagnoses: physician reexamination of the lay-interviewed cases in the general population. Arch Gen Psychiatry 1985;**42**:657–666.
88. Robins LN. Epidemiology: reflections on testing the validity of psychiatric interviews. Arch Gen Psychiatry 1985;**42**:918–924.
89. Eaton WW, Neufeld K, Chen LS, Cai G. Comparison of self-report and clinical diagnostic interviews for depression: Diagnostic Interview Schedule and Schedules for Clinical Assessment in Neuropsychiatry in the Baltimore Epidemiologic Catchment Area Follow-up. Arch Gen Psychiatry 2000;**57**:217–222.
90. Wing JK, Babor T, Brugha T et al. SCAN: Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry 1990;**47**:589–593.
91. Helzer JE, Spitznagel EL, McEvoy L. The predictive validity of lay Diagnostic Interview Schedule diagnoses in the general population. Arch Gen Psychiatry 1987;**44**:1069–1077.
92. Üstün TB, Kessler RC. Global burden of depressive disorders: the issue of duration. Br J Psychiatry 2002;**181**:181–183.
93. Spijker J, deGraaf R, Nolen WA. Determinants of poor 1-year outcome of DSM-III-R major depression in the general population: results of the Netherlands Mental Health Survey and Incidence Study (NEMESIS). Acta Psychiatr Scand 2001;**103**:122–130.
94. Parker G, Hadzi-Pavlovic D. Is any female preponderance in depression secondary to a primary female preponderance in anxiety disorders? Acta Psychiatr Scand 2001;**103**:252–256.
95. Stein MB, Fuetsch M, Müller N, Höfler M, Lieb R, Wittchen HU. Social anxiety disorder and the risk of depression: a prospective community study of adolescents and young adults. Arch Gen Psychiatry 2001;**58**:251–256.
96. Parker G, Wilhelm K, Mitchell P, Austin MP, Roussos J, Gladstone G. The influence of anxiety as a risk to early onset major depression. J Affect Disord 1999;**52**:11–17.
97. Wittchen HU, Hoyer J, Friis R. Generalized anxiety disorder: a risk factor for depression? Int J Methods Psychiatr Res 2001;**10**:52–57.
98. Breslau N, Schultz L, Peterson E. Sex differences in depression: a role for preexisting anxiety. Psychiatry Res 1995;**58**:1–12.
99. Murphy JM. What happens to depressed men? Harv Rev Psychiatry 1995;**3**:47–49.
100. Keller MB, Shapiro RW. "Double depression": superimposition of acute depressive episodes on chronic depressive disorders. Am J Psychiatry 1982;**139**:438–442.
101. Veroff J, Douvan E, Kulka RA. The inner American: a self-portrait from 1957 to 1976. New York: Basic Books, 1981.
102. Veroff J, Kulka RA, Douvan E. Mental health in America: patterns of help-seeking from 1957–1976. New York: Basic Books, 1981.
103. Eaton WW, Anthony JC, Gallo J et al. Natural history of Diagnostic Interview Schedule/DSM-IV major depression: the Baltimore Epidemiologic Catchment Area Follow-up. Arch Gen Psychiatry 1997;**54**:993–999.
104. Lehtinen JV, Lindholm T, Moring J, Puukka P, Vaisanen E. Incidence of mental disorders in the Finnish UKKI Study. Br J Psychiatry 1996;**168**:672–678.
105. Pakulski J, Waters M. The death of class. London: Sage, 1996.
106. Grusky DB, Weeden KA. Decomposition without death; a research agenda for a new class analysis. Acta Sociologica 2001;**44**:203–218.
107. Putnam RD. Bowling alone: the collapse and revival of American community. New York: Simon & Schuster, 2000.
108. Coleman JS. Foundations of social theory. Cambridge, MA: Harvard University Press, 1990.