

## Sex differences in depression

**Author:**

Wilhelm, Kathleen Anne

**Publication Date:**

1990

**DOI:**

<https://doi.org/10.26190/unsworks/11051>

**License:**

<https://creativecommons.org/licenses/by-nc-nd/3.0/au/>

Link to license to see what you are allowed to do with this resource.

Downloaded from <http://hdl.handle.net/1959.4/66164> in <https://unsworks.unsw.edu.au> on 2024-05-01

# **SEX DIFFERENCES IN DEPRESSION**

**K.A.Wilhelm**

**A thesis submitted for the degree of Doctor of Medicine, University of N.S.W.**

**December, 1990**

## **DECLARATION**

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which is a substantial extent has been accepted for the award of any other degree or diploma of a university or other institute of higher learning, except where due acknowledgement is made in the text.

## Acknowledgements

My first thanks go to the men and women in the cohort who made themselves available and gave up their time to be part of this study. It has been a pleasure to meet them all.

I would like to thank Professor Gordon Parker for the continued encouragement and supervision over more than a decade.

The N.S.W. Institute of Psychiatry provided a Research Fellowship which enabled the development of the Intimate Bond Measure and the five-year follow-up in 1983. The N.H.& M.R.C. provided the grant enabling Ms. Robyn Curtain to undertake the ten-year follow-up in 1988.

Robyn Curtain tackled the difficult job of relocating, interviewing the cohort members and generally co-ordinating the ten-year follow-up with her customary good humour, energy and enthusiasm.

Dusan Hadzi-Pavlovic for his thoughtful input, good humour, and patient answers to questions concerning statistical matters at many stages over this project. Kerrie Eysers, has provided great friendship and invaluable proof-reading, as have other friends who are acknowledged gratefully. Thanks also to my colleagues in the Mood Disorders Unit at Prince Henry Hospital, Drs. Philip Boyce, Henry Brodaty, Philip Mitchell and Ian Hickie for their interest and helpful comments.

Professor Gavin Andrews and his team were of great assistance in providing training and advice concerning the use of the DIS, but also wider interest in the project.

Cheerful secretarial support was provided by Joy Chaffey and Anne-Marie Rich. Noel Uhe was of great assistance in as a computer trouble-shooter and printer.



My husband, Peter Taylor, and children, Merry and Matthew have given me tremendous encouragement throughout, including accompanying me while travelling around N.S.W. over one school holiday in 1983 to meet teachers in rural areas. Thankyou, Peter for all the help with the final proof-reading.

I would like to dedicate this thesis to my father, Professor Don Wilhelm, who would have been delighted with the findings concerning the significance of paternal care.

"You agree then, I said, that men and women are to have a common way of life such as we have described - common education, common children; and they are to watch over the citizens in common whether abiding in the city or going out to war; they are to keep watch together, and to hunt together like dogs; and always and in all things, as far as they are able, women are to share with the men? And in so doing they will do what is best, and will not violate, but preserve the natural relation of the sexes."

Plato: 'Women As Equal To Men In The State',  
from The Republic, in The Dialogues of Plato

## CONTENTS

	Page
<b>Abstract</b> .....	1
<b>Chapter 1</b>	
<b>Definition of depression and issues of caseness</b>	
The concept of depression.....	3
Current theoretical classifications of depression.....	4
Primary and secondary depression.....	4
Unipolar and bipolar depression.....	5
Endogenous and neurotic depression.....	6
Depression categories in current diagnostic systems....	8
Diagnostic systems using operational criteria.....	10
Evaluation of changes brought about by use of operational criteria.....	15
Historical development of concepts of caseness.....	17
Use of self-report questionnaires.....	20
Use of semi-structured instruments.....	21
Discussion.....	24
Summary of themes and relevance to research question..	27
<b>Chapter 2</b>	
<b>Development of structured case-finding instruments</b>	
Structured case-finding instruments.....	29
Reliability and validity of structured case-finding instruments.....	33
The Present State Examination.....	33
The RDC/SADS System.....	34
The DIS System.....	37
Conclusions concerning properties of structured case-finding instruments.....	43
Selection of a case-finding instrument in the reported study.....	44
<b>Chapter 3</b>	
<b>Sex differences in the epidemiology of depression</b>	
Definition of terms.....	46
Review of sex differences in depressive experience....	47
Sex differences in rates of bipolar disorder.....	49
Sex differences in rates of unipolar depression.....	50
Review of general population studies using the PSE....	53
General population studies using the SADS/RDC system..	57
General population studies using the DIS.....	58
Review of mental health studies using a longitudinal design.....	61
Sex differences in rates of depressive symptoms.....	67
Data from Australian studies reporting rates of depression.....	68
Groups in which there have been a consistent lack of sex differences in depression.....	70
Summary.....	74
<b>Chapter 4</b>	
<b>Review of risk factors to depression with particular reference to sex differences</b>	
Examination of 'artifactual' factors.....	76
Design of instruments and possible sex biases in use of diagnostic criteria.....	76

Reporting of distress and depression.....	78
Help-seeking patterns.....	79
Examination of 'real' factors.....	80
Genetic factors.....	80
Biological factors.....	81
Psychosocial factors.....	84
Conclusions.....	93

## **Chapter 5**

### **The development of the Intimate Bond Measure**

Introduction.....	94
Consideration of existing measures.....	94
The development of the Intimate Bond Measure.....	98
Test construction and initial sample.....	99
Analysis of data from initial sample.....	100
Reliability studies.....	104
Validity studies.....	105
Discussion.....	106
Application of the IBM.....	109

## **Chapter 6**

### **Description of cohort study**

Introduction.....	111
Hypotheses.....	113
Sample selection.....	113
Entry assessment.....	115
Five-year review.....	116
Development of life event scale.....	118
Ten-year review.....	121
Evaluation of DIS data derived in mailed version.....	122
Tests of reliability of information.....	123
Use of the DIS to generate diagnostic categories.....	126
Special characteristics of the group.....	127
Analysis of data.....	128

## **Chapter 7**

### **Analysis of data in terms of sex differences**

Introduction.....	129
Psychosocial data at base-line, five-year and ten-year reviews.....	130
Perception of social support.....	135
Factors involved in choice of teaching as a career...136	
Experience of illness and help-seeking behaviour.....	139
Drug and alcohol use.....	144
Coping styles when depressed.....	146
Experience of life events.....	146
Self-reported experience of depression.....	151
Perception of self as a "nervous person".....	154
Self-report questionnaires.....	157
Self-report depressive symptom check-list.....	159
Summary of sex differences.....	159
Summary of absence of sex differences.....	161

## **Chapter 8**

### **Incidence and prevalence data for depressive disorders**

Introduction.....	163
Selection of diagnostic groupings.....	164
Use of the DIS.....	166

Rates for depressive disorders from 1983 and 1988	
DIS assessments.....	167
Rates for anxiety disorders from 1983 and 1988 DIS	
assessments.....	171
Relationship of anxiety to depressive disorders.....	173
Definition of caseness reconsidered.....	175
Rates for 'defined cases'.....	176
Summary.....	177

## **Chapter 9**

### **Reliability of data**

Introduction.....	179
Data from self-report measures.....	179
Data from DIS-derived diagnostic groupings.....	183
Data from 1978 assessment.....	184
Collaborative witness reports at 1983 assessment...	184
Inter-rater reliability between the two independent	
(1983 and 1988) raters.....	187
Comparison of rates for diagnostic groupings for	
periods prior to 1983, from 1983 and 1988 DIS data.	188
Comparison of reported numbers of symptoms and	
duration of episodes to 1983, from 1983 and 1988	
DIS data.....	200
Consideration of various methods of estimating	
lifetime case rates using 1983 and 1988	
DIS-generated data.....	204
Further examination of collaborative witness	
reports at 1983 interview to examine disparities	
in 1983 and 1988 DIS data.....	208
Data concerning recourse to help-seeking.....	211
General discussion of reliability of data.....	214
Summary.....	217

## **Chapter 10**

### **Risk factors predicting caseness**

Introduction.....	219
Choice of predictor variables.....	220
The position taken concerning relevance of negative	
life events.....	224
Method.....	224
Choice of dependent variables.....	225
Discriminant function analysis predicting caseness	
at 1983 and 1988.....	226
Importance of the risk factors in predicting	
caseness.....	236

### **Discussion**

The relationship shown for putative risk factors and	
depression.....	244
Neuroticism as a risk factor.....	245
Dependency as a risk factor.....	245
Low self-esteem as a risk factor.....	246
Perception of poor parenting as a risk factor.....	248
Feminine sex-role characteristics as a risk factor..	250
High trait depression as a risk factor.....	251
Absence of a 'core tie' as a risk factor.....	253
Female gender as a risk factor.....	253
Conclusions.....	254

## **Chapter 11**

### **Discussion of findings**

Introduction.....	256
<u>Issues pertaining to the cohort</u>	
How representative is this group of the general population?.....	257
Did those who volunteered for the study differ from their peers?.....	257
What were the special characteristics of the group?..	258
Were there particular issues related to teaching as a career?.....	258
Is there evidence that the males and females within the group differ from each other or expected social norms?.....	260
Are there sex differences in recall?.....	263
<u>Examination of sex differences</u>	
Socio-demographic variables.....	267
Are the lack of sex differences due to a Type II error?.....	268
Were there sex differences in perceptions of social support networks?.....	268
How do the rates for DIS-derived depression categories compare with other studies?.....	270
Why are lifetime prevalence rates higher in younger cohorts in recent times?.....	276
Have there been differences in social role expectations for younger cohorts?.....	278
<u>Other issues pertaining to the study</u>	
What is the relationship between anxiety and depression?.....	280
What is the relationship of life events to onset of depression?.....	282
How do the data from the study compare with the Bedford College model for social factors related to depression?.....	284
What is an appropriate definition of caseness?.....	288
How important was gender as a risk factor to depression for the caseness definitions used in this study?.....	290
Conclusions.....	290
Postscript.....	294
References.....	295
Appendix I The Intimate Bond Measure.....	
Appendix II Life events self-report scale.....	
Appendix III Results for discriminant function..... analyses reported in Chapter 10, with IBM scores.....	
Appendix IV Papers published on data from the thesis	

## ABSTRACT

The proposition that there is a female preponderance in depressive experience is examined in a cohort of young adults. A number of psychosocial variables (e.g. age, career opportunities, marital status) were initially controlled, with the expectation that there would initially be no sex differences and that a female preponderance would emerge over the subsequent decade as social heterogeneity occurred, allowing the identification of psychosocial risk factors to depression.

The classification of depression and the definition of caseness in non-clinical groups are reviewed, noting the application of structured case-finding instruments with particular reference to the Diagnostic Interview Schedule or DIS, which is used here. Literature concerning the epidemiology of sex differences in depression is considered in terms of real and artifactual explanations.

The study cohort commenced with 113 females and 56 males in 1978, with data on three assessments over ten years from 108 females and 53 males. The author conducted DIS interviews at a five-year follow-up then a lay interviewer conducted at a ten-year follow-up. There was moderate reliability over the ten-year period but evidence of some sex differences in recall of depressive episodes.

Three different definitions of caseness were used. Sex differences in rates of depression were not found at the five-year follow-up. Sex differences at ten years were found only for the caseness definition with the lowest symptom threshold.

There were no sex differences for self-report measures of trait, state depression

or perceptions of parental or current intimate relationships. Females rated higher on measures of neuroticism, dependency and femininity. Sex differences in social support and coping styles when depressed are described.

Using discriminant function analysis, the following risk factors relating to caseness were isolated: high trait depression, low self-esteem, high perceived maternal protection. Risk factors related to non-caseness included high paternal care and protection scores, and the presence of an intimate partner. A model linking perceived parental environment, the acquisition of self-esteem and consequent ability to sustain a good relationship with a partner and other social supports is considered. Gender did not constitute a significant risk factor, indicating the importance of social, rather than biological factors in determining onset of depressive disorders in young adults.



## CHAPTER ONE

### DEFINITION OF DEPRESSION AND ISSUES OF CASENESS

#### CONTENTS

The concept of depression  
 Current theoretical classifications of depression  
     Primary and secondary depression  
     Unipolar and bipolar depression  
     Endogenous and neurotic depression  
 Depression categories in current diagnostic systems  
 Diagnostic systems using operational criteria  
  
 Historical development of concepts of caseness  
     Use of self-report questionnaires  
     Use of semi-structured instruments  
 Discussion  
 Summary of themes and relevance to the current research objectives

The focus of this thesis is the examination of depression rates in young adults in a non-clinical cohort, to determine whether there are sex differences in rates of depression over time, with a further aim of identifying risk factors to depression. This chapter will discuss the concept of depression, its classification and definitions of caseness, with particular reference to non-clinical or general population settings.

#### The concept of depression

The term 'depression' conveys many meanings. Most commonly, it is used to describe a normal human emotion but it may also be used to convey an affect (the external manifestation of mood), a predicament (a state or being or condition that is unpleasant, trying or dangerous), a symptom (a complaint reported by a patient), a syndrome or disease (a constellation of symptoms and signs implying an underlying pathological process), or an illness (the manifestation of disease in the social context).

The Macquarie Dictionary (1982) defines depression as "dejection of spirits, and a state of despondency characterised by feelings of inadequacy, lowered activity,

sadness and pessimism", which embraces a variety of emotions, such as sadness, misery, loss of usual buoyancy and a lack of self-esteem. Such transient feelings of depression are common and frequent in non-clinical populations (Parker, 1979) and generally transient or ephemeral. Klerman (1974) has suggested that such 'normal' depression may have an adaptive function, as seen most clearly in the response of a child to separation. Here, 'depressive' behaviours such as crying will act as a signal, alerting the care-giver to the child's needs. In adults, depression may also be adaptive. Sadness that is a response to a loss (such as bereavement) has a restorative function for the subject and also communicates the subject's distress to others. Such awareness by others may then be accompanied by the accordance of certain privileges and offers of social support aimed at assisting with the grief work.

The features common to normal and pathological depressions include feelings of depression, lowered self-esteem, feelings of helplessness, hopelessness and self-criticism (Klerman, 1980). The determining factors for pathological mood states include longer duration of the episode, greater severity, the presence of certain features (e.g. change in psychomotor activity, anhedonia, preoccupation with morbid thoughts which may or may not be delusional) and inability to function in one's normal social roles. Help-seeking behaviour is also used as a criterion of pathological depression, but is problematic as there are subjects with pathological mood states who do not seek help.

### Current theoretical classifications of depression

#### **Primary and secondary depression**

Robins and Guze (1972) differentiated between **primary depression**, in which the subject had previously been well or had only had prior episodes of an affective disorder, and **secondary depression**, in which the depressive episode occurred in the setting of (i.e. secondary to) another physical or psychiatric disorder. Although Guze

(1971) had earlier hypothesised that primary and secondary depressions might differ in prognosis and treatment response, such differences have not been confirmed (Weissman et al, 1977; Giles et al, 1987). The classification, however, has been used for research studies, to exclude cases of depression associated with other disorders.

Andreasen and Winokur (1979) suggested that primary depression could be further subdivided into 'familial pure depressive disease' (subjects with a first episode after the age of forty and with a family history of primary depression, but without alcoholism or sociopathy), 'sporadic depressive disease' (similarly subjects with no positive family history), and 'depressive spectrum disease' (subjects, mostly women, usually first episode under forty, and with a family history of alcoholism or sociopathy in a first degree relative). This is a concept that continues to be investigated, although largely by Winokur's group (Winokur, 1985) and raises further demarcation issues concerning the boundaries between depression, sociopathy and alcoholism.

### **Unipolar and bipolar depression**

For primary affective disorder, Leonhard et al (1962) proposed a further distinction between bipolar disorder (subjects having a history of manic and depressive episodes) and monopolar disorder (where there is a history of only manic or depressive episodes) based on family history studies. They noted that patients with bipolar disease had higher rates of psychosis and their relatives had higher rates of suicide than those with monopolar depression. Mania is characterised by a subjective sense of well-being, with elated mood, poor judgement and impulse control, over-activity, decreased sleep and pressure of thought and speech. Psychotic phenomena (e.g. grandiose delusions, auditory hallucinations) may occur but are not necessary for the diagnosis. Subsequently, the concept of 'monopolar depression' has been retained as 'unipolar depression', while those subjects who have only manic episodes have been subsumed

under the category of bipolar disorder (W.H.O., 1978; Spitzer et al, 1978; American Psychiatric Association, 1980; Boyd & Weissman, 1981; Perris, 1982).

### **Endogenous and neurotic depression**

Unipolar depression has been further subdivided into two dichotomous groups. One type is variably labelled as endogenous, autonomous or psychotic depression, and viewed as qualitatively different from the other type, variously named exogenous, reactive, neurotic, or characterological depression. Kiloh (Kiloh et al, 1972) has summarised the position: "Psychotic or endogenous depression is a condition...with an imputed genetic or biochemical basis, whilst so-called neurotic depression is a diffuse entity encompassing some of the ways in which the patient utilizes his defence mechanisms to cope with his own neuroticism and concurrent environmental stress."

The psychotic or endogenous type has also been labelled as 'melancholia' particularly in the last few years (American Psychiatric Association, 1980; Klerman, 1980). Characteristic features of melancholia include motor retardation, lack of reactivity to environmental stimuli, anhedonia, excessive guilt, anorexia, severe weight loss, diurnal mood variation, terminal insomnia, and at times, hallucinations and delusions (when the term 'psychotic' is added). Unfortunately, the terms 'psychotic' and 'endogenous' have, at times, been used interchangeably, with many British writers using the term 'psychotic' to denote severity of depression (Carney et al, 1965), rather than the presence of specific psychotic symptoms. Until the 1970s, American psychiatrists tended to contrast 'neurotic' and 'psychotic' depressions in terms of severity, and 'reactive' and 'endogenous' depressions in terms of presumed aetiological factors (Gilbert, 1984), but since then have tended to use the term 'psychotic depression' to denote the presence of psychotic symptoms and have generally ceased using the terms 'reactive' and 'endogenous' depression (American Psychiatric Association, 1980;

Klerman, 1980).

The second type, neurotic or reactive depression is characterised partly by the lack of features characterising endogenous depression and is associated with reactivity of mood to environmental stimuli, initial insomnia, self-pity or doubt rather than guilt and anxiety symptoms (W.H.O, 1978; Klerman, 1980). The onset of such depressive episodes is thought to have a greater association with psychosocial stressors or conflicts than for endogenous depression.

There has also been a continuing debate as to whether the two types of depression (i.e. endogenous/psychotic and neurotic/reactive) simply reflect different dimensions of severity in one disorder (unitary view) or are separate entities (binary view). The terms 'endogenous' and 'reactive', which may refer both to type and aetiology, have been judged as unsatisfactory because more recent evidence belies the assumption that life events are only associated with reactive depressions and constitutional factors only with endogenous depressions (Paykel, 1974; Bebbington et al, 1988).

The Newcastle school attempted to distinguish between the neurotic and endogenous depressions by undertaking a discriminant function analysis of the scores of a heterogeneous group of depressed patients, and showing a bimodal distribution (Kiloh and Garside, 1963), later replicated by Carney et al (1965). Other groups have failed to replicate this finding (Kendell, 1969; Ni Bhrólchain et al, 1979). Kendell (1976) postulated a continuous spectrum of depressive disorders, or a psychotic-neurotic continuum, with a minority of patients at each end of the spectrum showing classical endogenous or neurotic disorders and the majority having a mixture of the two and lying along the continuum.

Akiskal et al (1978) noted that the term 'neurotic depression' is used in several different ways: denoting the absence of any psychotic phenomena, or a mild depression without vegetative symptoms, or the coexistence of neurotic symptoms such as anxiety, phobias and depersonalisation, or as a reaction to understandable life events, or as a characterological depression where subjects habitually react to normal or trivial stresses by developing depression because of underlying personality vulnerability. He argued that the lack of consensus over meanings of 'neurotic depression' indicated that the term should be dropped.

Wolpe contended that Akiskal's alternative categories of 'secondary dysphoria' and 'characterological depression' are only new names for identifiable categories of neurotic depression and argued for the retention of a positive diagnosis of neurotic depression (Wolpe, 1986). He argued that learned maladaptive anxiety responses are the basis of neurosis, and that well-defined patterns of neurotic depression are secondary to such anxiety. However, neurotic depression may be maladaptive and associated with anxiety without necessarily being secondary to it. Torgersen (1988) acknowledged that the term 'neurotic depression' represents a heterogeneous group of disorders, but stated that the DSM-III diagnostic categories are equally heterogeneous. He concluded that "perhaps the concept of minor depression used in RDC (*i.e. Research Diagnostic Criteria, Spitzer et al, 1978*) better represents the non-psychotic, non-melancholic major depression".

#### Depression categories in current diagnostic systems

Current diagnostic systems in clinical and research use are the International Classification of Diseases, now in its ninth edition (ICD-9), originating in U.K. and Europe under the auspices of the W.H.O. (W.H.O., 1978) and the Diagnostic and

Statistical Manual, with both third (DSM-III) and revised third editions (DSM-III-R), originating in the USA (American Psychiatric Association, 1980; 1988). There is now a fundamental difference in these systems, in that the ICD-9 provides broad and relatively non-specific definitions for psychiatric disorders, some of which are compromises between various European interests (Stengel, 1959). DSM-I (1952) and DSM-II (1968) were similar in concept to ICD-9, both using the neurotic/endogenous depressive categories in the spirit of the definitions supplied by Kiloh et al (1972) earlier and given for ICD-9 below. By contrast, the third edition of the Diagnostic Statistical Manual, DSM-III made a radical departure in providing definitions of disorders based on operational criteria (symptoms and signs) with cut-offs for inclusion in a diagnostic category and minimal reference to presumed aetiology. It is a multiaxial classification with the diagnosis of psychiatric disorder (i.e. symptom diagnosis) located on the first axis and four further axes describing personality, any concurrent medical conditions, predisposing life events and optimal level of function in the preceding year.

The latest version of the ICD system, ICD-9 endogenous depression (manic-depressive psychosis, depressed type) is defined as "an affective psychosis in which there is a widespread mood of gloom and wretchedness with some degree of anxiety" while neurotic depression is defined as a "neurotic disorder characterised by disproportionate depression which has usually recognizably ensued on a distressing experience.....there is often preoccupation with the psychic trauma which preceded the illness". A further category, 'adjustment reaction' covers "mild or transient disorders lasting longer than acute stress reactions...often relatively circumscribed or situation-specific, generally reversible"; these may be "brief", which includes grief reactions or "prolonged", lasting up to a few months. The term 'adjustment reaction' implies an understandable reaction to a specific stressor, while neurotic depression stipulates a level of depression disproportionate to the presumed stressor. There is also a category

'depressive disorder, not elsewhere classified' for "states of depression, usually of moderate but occasionally of marked intensity, which have no specifically manic-depressive or other psychotic features and which do not appear to be associated with stressful events or other features specified under neurotic depression".

#### Diagnostic systems using operational criteria

In this section, diagnostic categories for depressive disorders in DSM-III, DSM-III-R and Research Diagnostic Criteria or RDC will be described and compared.

The Research Diagnostic Criteria or RDC (Spitzer et al, 1978) were developed from the earlier Feighner diagnostic system (Feighner et al, 1972), with both relying on strict operationalised criteria. The Feighner system allowed 14 diagnostic categories, the categories of relevance to depression being primary and secondary affective disorder, with mania being included with primary affective disorder. In RDC, primary depression was divided into major (unipolar) depression and bipolar depression (subjects had also experienced manic episodes) and two new categories, minor depression and intermittent depressive disorder were created. These categories were intended to afford a broad coverage of depressive experience and to encompass the endogenous and neurotic depression categories which had been discarded.

As the RDC system was intended for use in clinical and non-clinical situations, allocation to RDC categories also requires the imposition of functional impairment criteria (seeking professional help, taking medication for the episode, or subjective judgement of a 'significant impact on life' because of the episode) which are not required for allocation to DSM-III categories. The DSM-III criteria were intended for clinical, rather than primarily for research use with the use of the term 'disorder' intended to imply that the episode is clinically significant. The reader is referred to a



paper comparing DSM-III and RDC criteria if more specific information is required (Williams & Spitzer, 1982).

The DSM-III system firstly defines 'disorder' (American Psychiatric Association, 1980) as "a clinically significant behavioral or psychological syndrome or pattern that occurs in an individual and that is typically associated with a painful symptom (distress) or impairment in one or more important areas of functioning (disability)" (p.6) implying a departure from a normal process or reaction.

In DSM-III, depressive disorders are included in the category of "affective disorders", where the essential feature is a sustained mood disturbance accompanied by related symptoms. There are five categories for depressive disorders: (1) bipolar depression (where there is current or previous evidence of manic episodes), (2) major depression with or without melancholia, (3) dysthymic disorder, (4) adjustment disorder with depressed mood (see Table 1.1 for comparison of these categories) and (5) atypical depression (depressive experience, unclassifiable elsewhere). The study to be reported uses both DSM-III and RDC depressive categories. Table 1.1 summarises criteria for duration, symptom numbers and impairment criteria for the DSM-III categories that are being discussed and allows comparison with RDC categories, from which they are derived (Spitzer et al, 1978; Williams and Spitzer, 1982).

DSM-III "major depressive episode" is defined as evidencing a "dysphoric mood or loss of interest or pleasure in all or almost all usual activities and pastimes", present for at least two weeks and accompanied by at least four of eight symptoms. These are effectively broad symptom groups, with all sleep disorders placed together, and with no distinction between initial and terminal insomnia, nor between appetite loss and appetite

Table 1.1 Criteria for depressive disorders

	Minimal duration	Number of symptoms required	Impaired functioning criteria
ICD-9 diagnoses			
Endogenous depression	NS	None	No
Neurotic depression	NS	None	No
RDC diagnoses			
Major depression			
- probable	1-2 weeks	4/8	Yes
- definite	2 weeks	5/8	Yes
Minor depression			
- probable	1 week		
- definite	2 weeks	2/16	Yes
Intermittent depression	2 years	2/16	Yes
	"much of the time"		
DSM-III diagnosis			
Major depression			
- total	2 weeks	4/8	No
- DIS/significant	2 weeks	4/8	Yes*
- with melancholia	2 weeks	additional	
- with delusions	2 weeks	criteria	
Adjustment disorder	NS	NS	No
Dysthymia	2 years	3/13	No*
	"relatively persistent"		
DSM-III-R diagnosis			
Major depression			
- total	2 weeks	5/9	No
Dysthymia	2 years	2/6	No*
	"more days than not"		
Zurich Study			
Extensive depressive episode			
EDE (SYM)	2 weeks	3/8 (males)	Yes
		5/8 (females)	Yes
EDE (WORK)	2 weeks	NS	Yes
General Health	"over past	5-7/30	
Questionnaire (GHQ)	few weeks"	(usually)	No
NS Not specified			
* RDC help-seeking/impairment criteria imposed with use of DIS			

\*\* Specifies subject as a case rather than defining diagnostic category

gain. There are additional requirements for "melancholia" and the presence of mood-congruent or mood-incongruent delusions (Table 1.1).

DSM-III major depressive episode and RDC major depressive disorder (probable) use the same duration and symptom criteria, while the report of one extra symptom is required for RDC major depressive disorder (definite). In DSM-III-R, major depressive episode requires 5 out of 9 symptoms present for the same two-week period, but one symptom has to be either depressed mood or loss of interest or pleasure. Eight of the nine symptoms are identical with DSM-III and the ninth is the depression item that was mandatory in DSM-III.

DSM-III dysthymic disorder requires the presence "most of the time" for two years of at least three of thirteen possible "symptoms characteristic of a depressive disorder but that are not of sufficient severity and duration to meet the criteria for major depressive episode" (APA, 1980). For DSM-III-R dysthymia, a depressed mood (present "more days than not") must be accompanied by 2 of a possible 6 symptoms. In DSM-III, it is possible to have a major depressive disorder superimposed on an episode of dysthymia, while DSM-III-R precludes this. RDC intermittent minor depression has the same minimum duration of two years as dysthymia but has subtle changes of wording in describing the minimum amount of time the subject should experience depression during the episode (Table 1.1), which is marginally less in RDC.

RDC has a category of minor depression which requires a minimum duration of one week for 'probable' and two weeks for 'definite' and a minimum 2 out of 16 possible symptoms. This is not equivalent to DSM-III adjustment disorder, which is defined as "a maladaptive reaction occurring within three months of a psychosocial stressor" (American Psychiatric Association, 1980). The maladaptive nature is indicated by

"impairment of social or occupational functioning" or symptoms "in excess of normal or expectable reaction". There are no minimum thresholds for symptom numbers or duration of episodes and complete resolution is assumed. The definition is the same in DSM-III-R but a maximum duration of six months is given. The concept of adjustment disorder is very different in its definition and emphasis compared to other DSM-III and RDC categories, and places more weight on subjective and rater interpretation (Table 1.1).

From Table 1.1 it can be seen that the ICD-9 system relies on clinical judgement to establish entry to diagnostic categories, whereas RDC and DSM-III diagnostic categories have strict entry criteria based on a minimum number and duration of symptoms. The RDC allows for the imposition of impairment criteria to define significant depressive episodes and these criteria have been applied to DSM-III diagnoses in the Diagnostic Interview Schedule or DIS (Robins, 1982) the instrument that is used in this study. The imposition of impairment criteria has an advantage in community studies where depressed subjects may not see themselves as depressed or requiring treatment.

Thus the category 'major depression' (using DSM-III or RDC systems) has strict entry criteria which may increase reliability, but it is a heterogeneous category including most of those episodes that would be called 'endogenous' and some that would have been labelled 'neurotic' in ICD-9 or DSM-II systems. The other 'neurotic' depressions now fall within the categories of dysthymia, adjustment disorder, cyclothymic personality or within the anxiety disorders. Winokur (1985a) was an early critic of the DSM-III classification. He stated that depression is a syndrome rather than an illness, which is "defined in DSM-III as a lowest common denominator....DSM-III essentially regards all depressions as equal and attempts to separate them with a

separate axis, ie personality disorders". He advocated the retention of the term 'neurotic-reactive' and presented criteria for the latter based on stormy lifestyle, personality assessment, family history and poor response to previous treatment.

#### Evaluation of changes brought about by use of operational criteria

The introduction of operationalised criteria in both the RDC and DSM-III systems was prompted by concerns about reliability of psychiatric diagnosis largely generated by researchers. Two studies, the United States-United Kingdom Diagnostic Project (Cooper et al, 1972) and the International Pilot Study of Schizophrenia (W.H.O., 1973) had drawn attention to lack of reliability of diagnosis in areas of depression and schizophrenia, with ensuing difficulties in international, even inter-regional, comparisons for clinical or research purposes. The impetus towards definition of reliable diagnostic and case-finding systems has been generated particularly by researchers interested in quantifying the rates of psychiatric disorder in a variety of populations, whether to look at changes in rates, risk factors determining rates or to assist health care utilisation practices.

Carroll (1989) discussed concepts of reliability and validity in terms of defining categories of depression against which the Dexamethasone Suppression Test could be evaluated. He stated that "when patients are recruited solely on the basis of 'operationally defined' clinical signs and symptoms, the acknowledged heterogeneity of the major depressive syndrome is left uncontrolled" and that "ICD-9 is generally regarded in the U.S. as inferior to DSM-III because ICD-9 diagnostic guide-lines are not operationalised". He considered that one test of diagnostic validity would be the demonstration of discrimination of different diagnostic categories by laboratory markers or differing treatment responses and pointed to "the disturbing fact that the current 'operational' diagnostic criteria (i.e. DSM-III categories) can yield groups of 'major

depressed" patients in whom it is impossible to demonstrate the therapeutic superiority of imipramine over placebo treatment!" while such discrimination is possible when ICD-9 and DSM-II categories are used. Thus, while the strictly operationalised diagnoses with less emphasis on aetiology and subjective distress (e.g. DSM-III diagnosis of major depressive disorder) have allowed for greater reliability in assigning diagnoses, the constructs may not have clinical validity.

Torgersen (1988) acknowledged that the term 'neurotic depression' combines a heterogeneous group of disorders but stated that the DSM-III diagnostic categories are equally heterogeneous and concluded that "perhaps the concept of minor depression used in RDC better represents the non-psychotic, non-melancholic major depression". His comments reflect the question raised earlier in the chapter by Carroll and Winokur as to whether the DSM-III categories actually constitute an advance in the conceptualisation of depressive categories, particularly for categories other than major depression (with or without melancholia).

Kendell (1989) commented that medicine has traditionally tied concepts of validity with a greater understanding of aetiology and underlying mechanisms (which could then be modified by treatment) and discussed what clinicians can do, by exercising their own clinical skills, to increase validity. He identified six validators of clinical syndromes: (i) identification and description of the syndrome, (ii) demonstration of boundaries between that and related syndromes, using statistical techniques such as discriminant function analysis and latent class analysis, (iii) follow-up studies, (iv) therapeutic trials, (v) family studies and (vi) association with fundamental abnormalities, either biological or psychological. He pointed to the following research strategies as being potentially useful: prospective follow-up studies based on serial interviews, therapeutic trials and family studies, both involving a broad

spectrum of diagnostic categories, and twin studies where alternative definitions of syndromes are used.

This study will use the first research strategy, a prospective follow-up with serial follow-up assessments, which is appropriate for a naturalistic study where no treatment intervention is considered but where there is an attempt to isolate predisposing social and psychological factors. The DIS was used to generate depressive categories using operationalised criteria for RDC and DSM-III systems and the instrument will be discussed in Chapter 2. While this review has implied criticism of the operationalised approaches of RDC and DSM-III systems, many of these short-comings have only become apparent in the past few years and this study commenced in 1978, when enthusiasm was high and short-comings less obvious. It is also important to use these 'newer' diagnostic categories to ascertain whether the atheoretical approach to diagnostic classification has heuristic value. One of the aims of the study was to examine the long-term reliability of depressive categories, while another was to examine risk factors to depressive disorder longitudinally. If risk factors were isolated in such a design, a case for validity of such depressive categories would be supported.

As the study to be reported involves a non-clinical cohort, the next section will consider definitions of caseness with particular reference to general population groups.

#### Historical development of concepts of caseness

From the 1940s to 1960s, psychiatric researchers used the clinician as the case-finding 'instrument'. Most studies at that time were disease-orientated, involving inpatient or outpatient groups. As the subjects had generally already presented for treatment, the concept of 'caseness' was not fully appreciated and generally equated with treatment presentation.

The Lundby study in Sweden (Essen-Moller, 1956; Hagnell, 1966) commenced in 1947, used a clinical interview as the case-finding instrument, and with the aim of evaluating the incidence and prevalence of anxiety and depression in an entire community of 2,550 people over a 25-year period. Essen-Moller stated that the study sought "to attempt some sort of description of all inhabitants, beyond those exhibiting conspicuous mental disease and abnormality...this type of an approach to the 'natural history' of personality was motivated by the conviction that mental differentiation is accomplished, not by the influence of human relations exclusively, but by an interaction of such influence with basic individual differences biologic in origin" (Hagnell, 1982). Essen-Moller and his team personally interviewed 98.8% of the inhabitants of Lundby using an interview of their own design. They supplemented their information with data from other informants and case histories, where appropriate. They also used team discussion to arrive at decisions concerning caseness.

The first large scale American epidemiological investigations of psychological health were the Stirling County study (Leighton, 1959) commencing in 1948 and involving 1003 people, and the Midtown Manhattan study (Srole et al, 1978) commencing in 1952 with 1,660 people. Both studies used symptom checklists which were administered by trained lay interviewers, rather than by psychiatrists. These checklists generated a score on a single dimension from health to pathology rather than assigning patients to diagnostic categories for treatment. This approach reflected the prevailing view that mental illness differed in degree rather than in kind, and was radical at the time when psychiatric opinion held that epidemiological methods would violate the concept of individuality and that for psychiatry "the most, if not the only, important research in this field are intensive studies of the individual, or at most, very small groups of individuals" (Lemkau, 1955). At that time, psychiatric opinion also



deemed much of the material used in psychiatric assessment to be out of conscious awareness and therefore not available simply by a process of direct questioning.

Like the Lundby study, the Stirling County study set out to determine the prevalence of anxiety and depressive symptoms in the community, using the Health Opinion Survey (HOS) and the Typology of Need for Psychiatric Attention Scale (PSYATT) which also provided some assessment of impairment and role dysfunction. Murphy later joined the Stirling County study, and has subsequently developed a computer programme that could be applied to the original data set to identify episodes of anxiety and depression comparable with the Research Diagnostic Criteria (Murphy et al, 1985).

The authors of the Midtown Manhattan survey postulated that "socio-cultural conditions, in both their normative and deviant forms, operating in intrafamily and extrafamily settings during childhood and adulthood have measurable consequences reflected in the mental health differences to be observed within a population". They did consider whether they should undertake an intensive psychiatric study of a few individuals or use a less intensive method for screening a larger population, finally deciding on the latter. They devised a questionnaire, the Home Interview Survey, to be administered by trained lay interviewers with data then reviewed by two psychiatrists who rated subjects on a continuum from 'well' to 'incapacitated', based on a combination of presence of symptoms and interference with life adjustment.

These studies were among the first to demonstrate that there was a substantial group of people with psychiatric disorders who did not present for psychiatric treatment and researchers started to question the relationship between caseness and patienthood.

### Instruments used to define caseness

#### **Self-report questionnaires**

Goldberg (1972) further examined the question of caseness in non-psychiatric patients by investigating levels of psychological distress in general practice attenders. He developed a self-report measure, the General Health Questionnaire (GHQ), to determine psychological caseness in a broad sense, with again no intention of defining specific psychiatric diagnoses. Although originally designed for use in general practice groups it has gained wider acceptance as a measure of current caseness with defined cut-off levels.

Self-report scales for depressive states (e.g. Beck et al, 1961; Zung, 1965; Wilson, 1979) convey some estimate of caseness with cut-off points that may vary with the researchers' needs, but were not designed to generate diagnoses. These measures are designed to evaluate subjective experience and are generally used as measures of depression or as screening techniques to identify subjects worthy of further consideration. A later questionnaire, the Centre for Epidemiological Studies-Depression Scale or CES-D (Radloff, 1977) was developed specifically as a screening measure in community surveys and, recently, Zimmerman (1987) developed a self-report questionnaire (IDD) to diagnose DSM-III major depression.

The Global Assessment Scale (GAS) is the broadest measure of impairment incorporating symptoms and role impairment. The scale provides examples of impairment in decile increments across a range of functions and is intended to measure cross-sectional function and to quantify change. However the reliability of such an approach is called into question as there may be greater changes in one area of functioning than another and raters may place emphasis on different areas when evaluating a subject. It has been included in the revised edition of DSM-III (DSM-III-

R; 1988) as a method of quantifying level of function.

### **Semi-structured instruments**

Angst and Dobler-Mikola (1984a) in Zurich, Switzerland, undertook a study of sex differences in a group of 591 young adults. They paid a great deal of attention to the problem of case definition from the outset. They noted that no valid definition of caseness existed and that 'cases' and 'non-cases' were on a continuum with an arbitrary line dividing the two. They considered that definitions of caseness should be flexible and suggested two definitions for use in their study. They had pre-tested the Present State Examination or PSE (Wing et al, 1974) but found it unsuitable for their purposes in that it only enquired about symptoms in the previous month and paid no attention to social problems and their consequences. They designed their own instrument, the Structured Psychopathological Interview or SPIKE, as part of a semi-structured interview.

They based their definition of a depressive 'case' on the Research Diagnostic Criteria but suggested differential cutoffs for minimum number of symptoms for each sex: three for males, five for females (Angst and Dobler-Mikola, 1984c).

For definition of a psychiatric 'case', they gave social consequences a higher priority than symptoms so that, as with the RDC, a 'case' still had to fulfil a minimum of two weeks' duration, with social impairment primarily at work, and which included unpaid work such as home duties. Impairment included reduction in performance, secondary conflicts and absence or loss of job. Loss of job was given last place on a rank order as some groups, such as housewives and unemployed people, cannot lose their jobs.

For episodes of depression, their definition of a psychiatric case is similar to RDC major depression (definite and probable) plus minor depression (definite), with less attention to the number of symptoms and with more attention being paid to the details subsumed by RDC's enquiry as to the significance or impact of the episode on the subject's life. The Swiss group's approach is particularly useful in general population studies as there will be a body of people who are on the border between 'case' and 'non-case', and whose allocation to caseness is arbitrarily decided on the basis of their remembering or forgetting perhaps one symptom. Definitions of caseness used in the case-finding instruments have been previously tabulated (see Table 1.1).

From 1968, the Social Research Unit at Bedford College has been examining the social causation of depression in women in the London borough of Camberwell (Brown and Harris, 1978). They defined 'disease' as "a departure from normality which may require medical intervention" and 'distress' as a "natural, if unpleasant process which can be expected to abate with time, without outside help". They contrasted 'cases' (subjects with a disease state, requiring intervention) with 'borderline cases' (where symptoms are atypical, less frequent or less intense, implying distress rather than disease). Both 'cases' and 'borderline cases' were based on PSE/ICD-9 categories using the Index of Definition (Wing 1970) to define entry to diagnostic categories, with accompanying operationalised case descriptions to illustrate typical examples of each category. Finlay-Jones et al (1980) have provided a checklist to demonstrate the clinical criteria that underlie their concepts and considered that the Bedford College definition of caseness (in terms of symptom patterns) was similar to DSM-III major depression or RDC 'probable' major depression, and that borderline caseness was similar to RDC minor depression (although with a minimum threshold of one symptom rather than two).

They also discriminated between vulnerability factors and provoking agents (Brown and Harris, 1978) and found that the presence of vulnerability factors increased the risk of subjects (who would have otherwise been regarded as 'borderline cases') becoming 'cases' in the face of a provoking agent.

Like the Bedford College group, a research team in Edinburgh (Surtees et al, 1983) has also concentrated their research efforts exclusively on women. They have used a Psychiatric Assessment Scale (PAS) which incorporates questions from the PSE and the SADS to generate diagnostic categories for PSE/Index of Definition, Bedford College criteria and RDC, to allow comparison with the various diagnostic systems. They were able to demonstrate changes in prevalence rates depending on which diagnostic system was used, with the PSE/Index of Definition and RDC systems giving similar prevalence rates and the Bedford College system tending to give the highest rates.

Vaillant and Schnurr (1988) compared a variety of definitions of 'caseness' to examine the validity of the concept in relation to a 40-year study of a group of male college graduates. They used six models for cases and/or psychiatric impairment: two categorical definitions, (i) a retrospective judgement, combining evidence of suspected psychic distress with independent assessments of impairment (e.g. seeking professional help, college-based psychological assessment, measure of adult adjustment) and (ii) a post-hoc estimation of lifetime DSM-III disorders. The other models were dimensional, (iii) the same measure of adult adjustment treated as a continuous variable, (iv) the Global Assessment Scale (GAS) assessed by an independent rater, (v) the Health Sickness Rating Scale (HSRS), and (vi) a measure grouping defence mechanisms on a nine-point scale ranging from 'mature' to 'immature' (Vaillant, 1977).

The choice of definitions of caseness and predictor variables were inevitably influenced by the available data from a longitudinal study of Harvard undergraduates (Vaillant, 1977), chosen for their expected mental health. They found correlations of 0.25 to 0.50 for a series of predictor variables compared with any of the six definitions of caseness, and reported that the six measures of impairment "were almost identical in their prediction of adult adjustment to late midlife". This result suggested that the six models had some common ground and that the concept of caseness can be arrived at through a variety of means. They raised the issue that most psychiatric disorders can be seen in dimensional as well as categorical terms and stated that "it is often desirable to use both categorical and dimensional approaches". They concluded that "an individual becomes, or ceases to be, a case through a complex interaction between impairment, host and environment", a remark which indicates some sharing of Brown and Harris' appreciation of the relationship between environmental effects and caseness.

#### Discussion of caseness definitions in non-clinical groups

The preceding section has reviewed some case-finding techniques used for non-clinical groups, some of which are studies of the general population, others have concentrated on smaller, more socially homogeneous groups.

The concept of 'caseness' in clinical groups is aligned with that of 'patienthood'; defined by Kraupl-Taylor (1972) as a person who is abnormal by the standards of the population in regard to at least one of the following (i) that the person feels therapeutic concern for himself, (ii) that this concern is also felt by his social environment, (iii) that there is medical concern for him.

This definition of patienthood reflects the original disease-orientated approach dating from the 1950s and developed from the study of hospital patients. It overlooks

concepts of help-seeking and illness behaviour and, in doing so, many of the special problems of case definition in general population studies. The latter studies involve subjects who are not presenting themselves for treatment and may not even see themselves as having a disorder.

One could argue that if people do not fit into this definition of 'patienthood' that there is no real reason to identify them. In the book "What is a Case?", which is devoted to this subject, Copeland (1981) makes the statement that "one of the sources of the problem is the misconception of a case as a unified entity" and challenges the assumption that the 'case' definition is based on a clinician's decision that such a subject would not be out of place in a treatment facility.

In terms of psychiatric research, it is also relevant to determine the natural history and degree of impairment associated with more minor forms of disorder that have not presented for treatment at that stage and whether there is any difference in outcome in treated and untreated disorders.

However, the epidemiologist or social psychiatrist who wishes to evaluate the range of depressive experience within a population must have some parameters by which to determine which subjects will be categorised as cases, either in terms of disturbances in social roles or functional incapacity. Copeland (1981) stated that caseness is "a concept created for a purpose is only useful in so far as it serves that purpose".

Two definitions of 'caseness' are pertinent to the area of depressive disorders. The first would lead to a concept of patienthood and treatment, the sort of notions embodied in the diagnosis of depressive disorder, and in predicting treatment outcome,

degree of morbidity and economic cost to the community.

The second would embody that concept of inability to function in normal social roles at the usual level of functioning and would be of value in identifying those who may be at risk for further impairment, assessing the value of the medical model, evaluating social factors contributing to depression and the social cost of such disorders to the individual and the community.

There are particular problems in investigating depressive experience in non-clinical samples, namely: (i) these subjects do not necessarily complain of distress or present for treatment, so that one has to make a decision of what is meaningful. (ii) Depression may be less 'severe' (in terms of numbers or types of symptoms, or degree of impairment) or it may present differently (e.g., be perceived as existential distress rather than symptom-based disease). (iii) Researchers may not wish to be restricted to diagnostic categories but to include flexible parameters of depressive experience. DSM-III has a category 'atypical depression' which is a catch-all category for all types of depression which do not fall into the rigid categories defined, but the group then becomes too heterogeneous to be meaningful. Otherwise, researchers can rate numbers of symptoms in existing categories as dimensions rather than categories or redefine the parameters, as Angst (1984c) has done.

The DIS (the case-finding instrument used in this study) first generates the same RDC and DSM-III diagnostic categories whether or not the subject has been a patient. The DIS then imposes the RDC help-seeking criteria which combine help-seeking and treatment criteria with a question designed to assess social role impairment to consider those subjects who have not been patients. The next chapter will consider the properties of some semi-structured case-finding instruments in common use.



### Summary of themes and relevance to research question

Two themes have been developed. The first is to describe some of the difficulties in categorising depression and to highlight the definitions used in the two most commonly used current diagnostic systems (ICD-9 and DSM-III). The second is to examine the concept of caseness in depressive disorders. The definition of caseness used has been influenced by whether the subjects being investigated were from a clinical or non-clinical population, the availability of case-finding instruments and the research questions being asked.

From the 1950s to mid 1970s, there were sociological studies of general population groups that looked at concepts of distress and mental health (that were often poorly defined), while psychiatric studies looked at clinical samples using clinical criteria (that were also often poorly defined in terms of operational diagnostic criteria).

Prior to the 1970s, community-based studies used either clinical interviews or dimensional constructs (e.g. self-report symptom measures or measures based on a continuum from health to illness) to categorise depression. In the last fifteen years there has been interest in the rates of depressive diagnoses in the general population, coinciding with the development of more reliable diagnostic systems and related case-finding instruments which are based on operational decisions for generation of RDC or DSM-III categories.

The drive towards more reliable definitions of depressive categories has been noted, culminating in the operational RDC and DSM-III classification systems. The American Psychiatric Association proposed an atheoretical approach to classification

embodied in the DSM-III diagnostic classification. In relation to depression, this meant replacing categories of neurotic/reactive and endogenous/psychotic depression (still used in the ICD-9 classification) and with their implicit assumptions concerning causation, with categories of major depression and dysthymia where no such causal implications were drawn. The DSM-III system implies a unitary (dimensional) approach, with operationalised diagnostic categories, based on symptoms and signs, with little reference to presumed aetiology. However the category 'adjustment disorder' does call for the clinician to make some subjective judgements about the maladaptive nature of the reaction and whether the symptoms are in excess of a normal reaction to the stressor.

The British and Europeans have been more conservative and retained the ICD-9 system. This implies a binary approach based on diagnostic categories which are loosely defined in terms of symptoms, signs and presumed aetiology. The unitary/binary debate has continued in British and European circles.

While the categorical approach has proved useful in improving reliability, validity is largely unexplored. These issues will be discussed in the next chapter in relation to case-finding instruments based on RDC, DSM-III and ICD-9 diagnostic systems.

Any consideration of depressive disorder categories overlaps with the concept of caseness and the conclusion reached is that the categorical approach is useful and such an approach is used in this study. In general one should have flexible methods of determining caseness, determined by the research question that is being addressed.

## CHAPTER TWO

### DEVELOPMENT OF STRUCTURED CASE-FINDING INSTRUMENTS

#### CONTENTS

Structured case-finding instruments  
 Reliability and validity of structured case-finding instruments  
   The Present State Examination  
   The RDC/SADS system  
   The DIS system  
 Conclusions concerning properties of structured case-finding instruments  
 Selection of a case-finding instrument in the reported study

#### Introduction

The need to make reliable diagnostic decisions for psychiatric research purposes was recognised in the early 1970s. Robins (1989) later stated that the goal was "to construct standardised interviews that would function more like a psychiatrist does during a clinical assessment...that is, they would be tightly tied to official diagnostic criteria; they would endeavor to distinguish clinically significant symptoms from the trials and tribulations of everyday life...not to write an interview equivalent to the behavior of particular psychiatrists; instead ...to come as close as possible to what the ideal psychiatrist would do if he properly interpreted and rigorously followed the criteria in the diagnostic systems being assessed".

#### Structured Case-finding Instruments

The Present State Examination or PSE (Wing et al, 1974) was the first instrument to be developed using a semi-structured interview, with operational criteria for symptoms and signs present over the previous month. There are 140 items, each of which is rated on a 3 or 4 point scale. The interview is "basically...a check list which systematically covers all the phenomena likely to be considered during a present state examination". It is intended for use by a clinician who uses his judgement as to the relevance of the material presented to him, but a training programme is recommended, during which the interviewer learns the definitions of each item from a manual which includes a glossary of symptoms and signs.

The PSE was originally intended for use with hospital patients, but a 40-item version was subsequently developed for non-patient groups. Diagnostic categories based on the ICD-9 system can be made and the accompanying computer programme (CATEGO) generates an Index of Definition or ID score (Wing, 1976) for caseness (an ID score of 5 or more) and borderline caseness (an ID score of 3 or 4), but the instrument was not intended to give a diagnosis, rather a measure of 'present state'.

American psychiatrists also recognised the advantage of structured case-finding instruments. The Renard Diagnostic Interview or RDI was developed to elicit the symptoms required to generate the Feighner diagnostic criteria and was intended for the use by physicians and non-clinicians (Helzer et al, 1981).

The Schedule for Affective Disorders and Schizophrenia, or SADS (Spitzer et al, 1978), was designed as the primary diagnostic tool for the National Institute of Mental Health (NIMH) Collaborative Study of the Psychobiology of Depression (Katz et al, 1979). It is a structured interview schedule to be used by a clinician to generate the RDC categories (RDC depression categories were described in the previous chapter). The first section assesses symptom severity for the nadir of the current episode and for the preceding week. A second section deals with previous episodes of illness. The diagnostic decisions are then made by the interviewer referring to the Research Diagnostic Criteria. There are three versions, SADS, SADS-L (lifetime version) and SADS-C (where change in symptoms is being measured). All these instruments are intended for use only by clinician researchers and training is recommended. There is also a Family History-RDC (Andreasen et al, 1977) for use with members of the subject's family, if the subject is unavailable for direct evaluation. The SADS has been used in depression research because it identifies 10 subtypes of major depressive disorder found in the RDC, covering a broad range of depressive experience.

The Diagnostic Interview Schedule or DIS (Robins et al, 1981) was designed specifically for use in an ambitious epidemiological survey, the Epidemiological Catchment Area (ECA) studies, performed by a number of independent research teams at five USA centres, also under the auspices of the NIMH (Regier et al, 1984). The authors state that the "broad aims of the ECA program are the historical goals of psychiatric epidemiology: to estimate the incidence and prevalence of mental disorders; to search for etiological clues; and to aid in the planning of health care services and programs". They stated that the study was innovative in a number of ways, (i) by integrating surveys from community and institutional populations, (ii) by collecting prevalence and incidence data; (iii) by the use of multiple research sites with collaborative collection of data and (iv) by "field validation" of the DIS (Eaton et al, 1981). The SADS was not considered appropriate for the ECA study because it required decisions by clinicians (the ECA study design called for a reliable case-finding instrument for use by a large group of lay interviewers) and as it generated RDC categories rather than the DSM-III categories intended for use in the ECA study. Furthermore, the ECA researchers wished to examine prevalence over short periods of time (i.e. two weeks and six months) in addition to life-time and the SADS was not able to provide the short term prevalence data.

The DIS is a highly structured interview schedule using a probe flow chart which allows lay interviewers (after a one to two-week training period) to assess the significance of symptoms. It is based on the Renard Diagnostic Interview and, like the RDI, uses a probe system to determine impairment and to distinguish psychiatric symptoms from consequences of medical illness and the effects of drugs and alcohol. There is a computer algorithm that can be applied to generate Feighner, RDC and DSM-III diagnostic categories for symptoms over the entire span of the subject's life; also period prevalence data over the previous two weeks, month, 6 months and 12 months, as well as for the nadir of the current episode. The DIS was designed for use with psychiatric inpatient, outpatient and general population groups, and is the only measure that was specifically designed for use by lay interviewers with non-clinical groups. The DIS has achieved a fair degree of

success in the hands of lay interviewers as the wording of every question in the structured interview is strictly predetermined and precoded, as are the decision rules (presented as a Probe Flow Chart) used during the interview.

The DIS provides a broader coverage of psychiatric diagnoses than PSE or SADS and fewer subjects are relegated to residual categories. Its highly structured format has the advantages of brevity and the possible adaptation for use by telephone interview (Wells et al, 1988) or in a computer-assisted version (Blouin et al, 1988). There is also provision for collection of a total symptom count regardless of diagnostic category and allowance for multiple simultaneous diagnoses (e.g. the co-occurrence of anxiety and depressive disorders) if the hierarchical rankings of categories implicit in DSM-III are not imposed.

There is now an instrument called the CIDI or Composite International Diagnostic Interview (Robins et al, 1988) combining the DIS and PSE, which generates diagnoses for Feighner, RDC, DSM-III and DSM-III-R criteria, as well as certain CATEGO classes derived from ICD-9. This instrument is currently undergoing an international multi-centre reliability trial, for which no results are yet to hand (Sartorius, 1989).

The Structured Clinical Interview for DSM-III or SCID (Spitzer & Williams, 1983) was specifically designed by the authors of the SADS/RDC system to follow the decision rules used in making DSM-III diagnoses in much the same way as the SADS was designed to elicit RDC categories. Each section commences with the essential criteria for each diagnosis and then provides prompts and questions to determine whether the additional criteria are met. This measure is intended to be used for a current episode and to be administered by a clinician. It allows for incorporation of other relevant material, such as reports from clinical records and embodies the changes in DSM-III-R. It has broader coverage of diagnoses than the DIS and will generate data for other DSM-III axes. Rabkin and Klein (1987) noted the possibility of greater validity than the DIS due to the broader data base and more flexible system of enquiry. However the SCID is intended for use by

clinicians rather than lay interviewers, so that the DIS is likely to remain the instrument of choice in epidemiological surveys.

### Reliability and validity of structured case-finding instruments

#### **The Present State Examination**

The PSE is now in its ninth version. Several versions used much the same format for reliability studies for successive versions (Wing et al, 1974). Each reliability study consisted of interviews for 170 to 190 patients, approximately two-thirds of whom were interviewed simultaneously by two psychiatrists (to check inter-rater reliability), with the remainder interviewed on two separate occasions, usually one to two days apart (to check test-retest reliability). In an early reliability study, across all diagnoses, Kendell (1968) found a mean kappa value of 0.71 for inter-observer agreement and 0.41 for test-retest reliability. Intra-rater reliability was quoted in terms of product-moment reliability coefficients, which were in the range of 0.80 to 0.95 for depressive symptoms in both reliability studies (Wing et al, 1967; Kendell et al, 1968), with situational anxiety symptoms having the lowest values (0.58). Inter-rater reliability using the 40-item version with non-clinicians (Cooper et al, 1977) returned lower reliability coefficients for depressive symptoms, with a mean of 0.67. Wing's group (Wing et al, 1974) acknowledged the possibility of spuriously high reliability rates in their group as the the rating psychiatrists were all close colleagues who employed very similar clinical decision-making practices. Another study by the same group (Wing et al, 1974) examined differences between psychiatrists and inexperienced raters. The clinically experienced psychiatrists reached high inter-rater reliability within one week of training, while those with less experience firstly rated more, then fewer, positive items than the experienced raters. However, the conclusion was that the PSE training could provide an "acceptable degree of reliability and repeatability at all stages of the diagnostic process" (p 68).

In a study (Sturt, 1981) examining whether the PSE is a valid measure of current severity of psychiatric disorder, a group of 800 subjects from the general population in

Camberwell, U.K., was compared with three groups comprising (i) all patients attending the casualty department following deliberate overdose of medication, (ii) patients between 18 and 64 years (excluding schizophrenia and drug/alcohol abuse) attending a general psychiatric outpatient service, and (iii) a consecutive series of inpatients to a general hospital area-based service. Those subjects who were inpatients were noted to have different case rates using the Index of Definition (with higher total PSE scores), different symptoms and greater social impairment in inpatient groups (with some symptoms reported frequently for inpatients being rare in the outpatient groups). The authors viewed these data as demonstrating a relationship between patienthood, clinical severity and Index of Definition categories. This is despite the possibility (which was not addressed) that such psychosocial issues as employment, marital state and previous numbers of hospital admissions may also affect the decision to admit a patient to hospital.

There are limitations to the use of the PSE. Firstly, the measure was only designed to give current prevalence data. Secondly, as the interview was largely based on material derived from patients referred for psychiatric treatment (usually after admission to hospital), its relevance to case-finding in general population groups may be limited for a number of reasons: (i) symptoms may be less well defined, or (ii) have different thresholds of severity, or (iii) some of the symptoms defined in the PSE for inpatients may be less common or of little relevance in non-clinical groups. The authors suggest that screening measures may compensate for this. However, the PSE has continued to be used as a case-finding instrument, often preceded by the GHQ as a screening measure (Henderson et al, 1979) and some studies using this methodology will be mentioned in Chapter 3.

### **The RDC/SADS system**

In the development of RDC (Spitzer et al, 1978), there were two studies of inter-rater reliability which involved 218 psychiatric clinic inpatients and one of test-retest reliability (with a sub-group of 60 from the same group), as part of a NIMH project, the Collaborative Program on the Psychobiology of Depression. The kappa values for RDC



primary major depression in the two studies of inter-rater reliability ranged from 0.48 to 0.87 and for test-retest reliability, from 0.59 to 0.86.

The Schedule for Affective Disorders and Schizophrenia or SADS (Endicott & Spitzer, 1978) was developed as a case-finding instrument to generate RDC diagnoses, and the initial reliability studies involved a group of 150 hospital in-patients that were used as part of the same NIMH multi-centre collaborative project, the Collaborative Program on the Psychobiology of Depression. Patients with a diagnosis of mania or depression were interviewed separately, with interviews within 48 hours of each other. The authors reported correlations of over 0.60 for 90% of the items (when taken item by item) for inter-observer reliability, and for 82% of items for test-retest reliability. Correlations for test-retest reliability (on a sub-group of 60 patients) ranged from 0.67 for the presence of anxiety to 0.83 for 'endogenous features'. The Symptom Check List or SCL-90 (Derogatis et al, 1973) was also completed by 144 of the 150 subjects involved in the reliability study with correlations ranging from 0.68 between depression on SCL-90 and 'depressive mood and ideation' generated from SADS, to 0.47 for SCL-90 depression and 'suicidal ideation and behaviour' from SADS.

Mazure and Gershon (1979) examined test-retest reliability using SADS-L with a group of 49 patients, their first-degree relatives and medical controls, using two interviews about seven months apart. The test-retest reliability was "high" (kappa of 0.79) for a lifetime diagnosis of major depression.

There were no attempts at validity studies at that time but Leckman (Leckman et al, 1982) later used a 'best estimate' method utilising information from three different sources (two direct interviews with raters each being blind to the findings of the other, family history data, medical records) to study the reliability and validity of lifetime diagnosis. Most interviewers used SADS-L and all made RDC diagnoses. Reliability rates for 'best-estimate' rates of the two clinicians returned kappa values between 0.46 (for minor

depression) to 0.87 for hospitalised unipolar depressives. There was a kappa value of 0.80 for the category 'never mentally ill'.

Andreasen's group (Andreasen et al, 1981) examined test-retest reliability for the SADS-L over varying periods of time. Their first "short-term reliability" study involved interviews of the same patients on two occasions in one day. Reliability between morning and afternoon raters for RDC depressive subtypes ranged from kappa values of 0.19 (for incapacitated) to 0.87 (for primary major depression). There was also high reliability for help-seeking behaviour (kappa=0.64), social role impairment (kappa=0.73), number of symptoms ( $r=0.84$ ) and age of first episode ( $r=0.84$ ).

The second "longer term reliability" study called for a consensus diagnosis by guest raters using SADS-L generated, lifetime diagnosis data on subjects seen by the host research team in the previous six months. Kappa values for comparison of the two diagnoses, ranged from 0.20 (for RDC 'recurrent') to 0.75 for RDC primary major depression. The authors concluded that short-term reliability was high and longer-term reliability was "acceptable", allowing reliable long-term diagnosis in a nonpatient population. This is an optimistic conclusion to draw from these two studies, as the first measured test-retest reliability on the same day, which is not a rigorous test for an instrument (SADS-L) which was intended to measure lifetime disorder; and both these reliability studies were carried out on clinical groups and the results cannot necessarily be extrapolated to non-clinical subjects.

Bromet et al (1986) investigated test-retest reliability over an eighteen-month period, using the SADS-L in a population of 391 women selected for observation of long-term effects on mental health of experience of a nuclear power disaster. Reliability was found to be poor, with only 38% of the women reporting RDC episodes of major depression on both occasions. Reliability over the eighteen month period was influenced by a number of clinical status factors. Those who had suffered further episodes of

depression were more likely to recall earlier episodes and the women who were able to recall episodes of major depression also recalled help-seeking behaviour, medication use and age of first episode, but were not reliable at recalling length of longest episode or total number of symptoms. The authors stated that this non-clinical group returned lower rates for reliability than patient samples. They noted that their test-retest period is longer than other studies and question whether the SADS-L itself may have contributed to unreliability, and added "it is conceivable that a fully structured instrument, such as the DIS, might increase test-retest reliability". They alluded to the fact that a non-clinical group does not have as many prompts to recall (e.g. taking medication, visits to doctor) as a clinical group. However they did not mention the special characteristics of the group under review (after a nuclear power station disaster) in that the experience of the disaster may have been sufficient to override expected levels of recall of depressive experience.

### **The DIS System**

There was only one attempt to evaluate the validity of the DIS prior to commencing the ECA study (Robins, 1981). A series of 216 patients was interviewed twice, once by a psychiatrist and once by a lay interviewer. All of the interviewers were inexperienced - the psychiatrists had all only recently completed their psychiatric training and other interviewers were college graduates with no previous interviewing experience. It is not made clear whether this was a specific decision not to use experienced clinicians (who would presumably bring more of their own clinical experience to bear on the decision making), or whether it was fuelled by expediency. The diagnosis made by the psychiatrist was used as the yardstick, with kappas ranging from 0.63 to 0.68 for agreement across the various depressive disorders. The authors argue that these interviews represent a study of criterion validity, "if one considers the DIS in the hands of a psychiatrist a criterion" (Eaton et al, 1981). However, their procedure actually assessed inter-rater reliability, not validity.

Three (New Haven, St. Louis and Baltimore) of the five centres in the multi-centre study undertook further reliability studies (which they termed "field validity studies")

simultaneously with the data collection for the ECA study. Each centre used a different method for assessing the performance of the initial DIS interview in the hands of the lay interviewers.

The Baltimore group (Anthony et al, 1985) had used the GHQ as an separate measure of current psychiatric distress. They recalled 392 subjects who were 'positive' in terms of DIS- generated diagnoses and GHQ caseness to be re-examined by one of four recently qualified psychiatrists. Three-quarters of the subjects were re-interviewed within three to four weeks of the original interview, and all within the next three months. The psychiatrists conducted a standardised clinical interview which included a checklist of DSM-III items and the full PSE. They were encouraged to ask about the subject's developmental history, health and personality issues; in essence, to cover material generally accessible to a clinician, but not available to lay interviewers using the DIS. The DSM-III checklist was thought necessary as clinicians generally do not adhere rigidly to DSM-III diagnostic decisions, even when they claim to be 'going by the book' (Jampala et al, 1987).

In the Baltimore study, the authors reported kappa values from -0.02 to 0.35, with a value of 0.25 for major depression, for which their claim of "moderate" agreement between the two raters is unsupported even for values at the higher end of the range of kappa values. The authors note some of the areas of possible discrepancy, which included inadequate information being given to one rater, difference in symptom pattern review on the two occasions of interview and incomplete coverage of disorders by the DIS (so that DIS-derived information is insufficient to fully discriminate between episodes of major depression and other affective disorders). The authors also noted that the DIS places a high degree of reliance on the ability of subjects to fully discriminate between symptoms due to substance abuse, physical problems and those due simply to psychiatric problems. They recommended that information be gathered from alternative sources rather than full reliance be placed on one cross-sectional diagnostic interview to generate reliable data.

The St. Louis group (Helzer et al, 1985) used nine psychiatrists with varying levels of experience to re-examine 394 subjects (within a few weeks of the original lay-administered DIS interview) using the DIS and a DSM-III checklist while being blind to the original diagnoses and caseness of the subjects. This study reported overall agreement of 0.89, with sensitivity of 0.63 and specificity of 0.99 for subjects with major depression. They noted that diagnostic agreement is likely to be lower in general population samples - where there are fewer symptoms, lesser severity and less likelihood of subjects presenting for treatment (in that a discussion with a clinician increases the likelihood of episodes being remembered and acts as a rehearsal for the DIS interview situation).

A German form of the DIS and a German instrument deriving ICD-8 diagnoses were used with 171 former psychiatric patients and 158 subjects as part of a seven-year prospective general population study (Wittchen et al, 1985) at two interview occasions within three days of each other. Interviews were conducted by one a team of eight "experienced physicians" (which implied at least two years psychiatric training after gaining a medical degree) and 12 psychologists. For inter-rater reliability, they reported kappa values ranging from 0.90 for current episodes of major depression to 0.52 for simple phobia, and "considerably lower" rates for dysthymia and panic disorder. They questioned whether the low reliability of rates for dysthymia was a function of the low base rate for the disorder. The diagnostic categories were then compared (where possible) to ICD-8 diagnoses, giving kappa values of 0.84 between ICD-8 unipolar affective psychosis and DSM-III major depression, and 0.81 with RDC major depression, and kappa values of 0.81 between ICD-8 depressive neurosis and DSM-III dysthymia, and 0.74 for RDC intermittent minor depression. They concluded that the DIS had "sufficiently high overall specificity and sensitivity as a case-finding instrument in the general population and "surprisingly high concordance of most DIS diagnostic classes and comparable ICD diagnoses".

Helzer et al (1987) questioned the most appropriate method of measuring validity of diagnoses generated by the lay DIS interview. They carried out a twelve-month follow-up

of the group of 370 subjects used in the comparison of clinician and lay DIS interviews previously noted (Helzer et al, 1985) and also examined family history results to compare the predictive power of lay interviewers' and psychiatrists' index diagnoses. The greatest difficulty was found with cases that were originally near the threshold level, so that subsequent reporting of one fewer symptom may have determined presence or absence of a disorder at the second interview. They found that the lay interview compared favourably with clinician checklist diagnoses in predicting diagnostic consistency and psychiatric diagnosis within the patient's family. Like Bromet, they found "a considerable fall-off in lifetime diagnoses from wave I to wave II".

A later study of test-retest reliability (Wittchen et al, 1989) involving 60 psychiatric in-patients used the DIS (version 3) on two occasions over a one to four-day period with two different raters (psychiatrist or psychologist). They found that subjects reported fewer symptoms at repeat interview for recent episodes but that there was no significant change in lifetime diagnosis rates. They interpreted this as a tendency for patients to deny symptoms in the present which they were willing to admit to in the past. However, this explanation does not account for the finding that the symptoms had been acknowledged at the initial interview. There was higher concordance for age of onset data for more 'severe' (termed "psychotic") disorders (using ICD-8 categories of schizophrenia, schizoaffective, unipolar and bipolar affective psychoses), than for less severe (termed "non-psychotic") disorders (using ICD-8 categories anxiety neurosis and depressive neurosis), which the authors considered could affect reliability of rates for less severe disorders seen in general population samples.

In a more recent study in Detroit, U.S.A. (McLeod et al, 1990), 353 (73%) of 484 subjects who had been interviewed by lay interviewers using the DIS in a community sample were re-interviewed by a team of clinicians (who were all psychiatric social workers). The median lag time to second interview was 11 weeks. The DIS interviews generated RDC diagnoses for individual and combined depressive (RDC case) categories

for the six-month period prior to the initial interview. The authors reported kappas of 0.39 for all RDC cases, 0.28 for major depression, 0.10 for intermittent minor and 0.08 for minor depression. They noted that longer, more severe episodes were more consistently recalled and that inconsistency in reporting was most often associated with the timing of the episode (i.e. whether or not the episode has occurred within the last six months). They considered such issues as differences in empathy between clinical and lay interviewers, slight differences in wording of the questions at each interview and cue effects from the original interview, but concluded with the opinion that the major cause of differences was inconsistent recall by the subjects.

Klerman (1985a) has reviewed the use of the DIS in the Epidemiological Catchment Area (or ECA) study (Eaton et al, 1984). He noted that earlier surveys of psychiatric epidemiology had relied on measures of impairment rather than clinical diagnosis and commented on the increasing use of lay interviewers in epidemiological studies, clinical studies and therapeutic trials before the question of the reliability of lay interviewers versus psychiatrist interviewers had been resolved. He stated that "in retrospect, it would have been desirable if more extensive evidence as to the reliability and validity of the DIS and DSM-III criteria were available" prior to commencing use of the DIS and regretted the reliance on the one-stage method used in the ECA study. He suggested for future studies the use of a two-stage method, with a screening measure to identify subjects at risk, who would then undergo a more intensive diagnostic process. Such screening designs should have high false positive rates to ensure that most likely 'cases' are included in the screening net.

Parker (1987) has questioned the low lifetime prevalence estimates reported in the ECA study and, by implication, the reliability and validity of the DIS-derived data. He concluded that there is "reasonably high" inter-rater reliability for current DIS-derived diagnosis of depressive illness in patient samples, but less inter-rater reliability when highly trained psychiatrists were compared to lay interviewers. He drew attention to the 'best

estimate' method used by Leckman (Leckman et al, 1982) in testing the validity of SADS-L measure and noted that no similar procedure had been used with the DIS.

Spitzer (1983) advocated use of a similar technique termed the Longitudinal, Expert, and All Data (or LEAD) standard which involved use of all available sources of data rather than total reliance on a cross-sectional interview from subject only. This approach was incorporated in the SCID interview which was cited earlier in the chapter.

Robins (1989) concluded a review of structured case-finding instruments by stating that "we presently have no better method for assessing psychiatric illness than the standardised interview. It may well be most accurate in the hands of a clinician, but it is not economically feasible to have all epidemiological studies carried out by psychiatrists, so long as lay interviewers do reasonably well".

Burke (1986) reviewed the performance of the DIS in categorising clinical and non-clinical groups and concluded that the DIS performed "adequately as a case-finding instrument for...alcohol disorders and depression; its performance in panic disorder seems questionable". He advocated further development of tests applying the Longitudinal, Expert and All Data (LEAD) standard (Spitzer et al, 1983) and considered that it was fruitful to look for areas of discrepancy between clinical interview and instruments like the DIS, which he listed as (i) the nature of the interview, (ii) allocation of marginal or borderline cases, (iii) subject variance over time (with consideration of how such variation could best be assessed), (iv) differing interpretation of items and exclusion rules over time, (v) misallocation into specific categories, (vi) determination of clustering of symptoms and timing of episodes, (vii) the inherent unreliability of assessments. He concluded that inter-rater reliability for the DIS was "adequate", that agreement between the DIS and a clinical diagnostic interview was "fair to good", and that discrepancies between the two were not necessarily due to "errors" in either assessment procedure.



### Conclusions concerning properties of structured case-finding instruments

The properties of the most commonly used structured case-finding instruments (PSE, SADS and DIS) have been considered. A range of studies using both clinical and non-clinical populations was reviewed. There are comparable levels of reliability for each measure for inter-rater reliability and test-retest reliability. There are variations for different diagnostic categories, with major depression generally having adequate inter-rater and test-retest reliability, and dysthymia and simple phobias being less reliable. Most of the psychiatric raters used in the reviewed studies have been trainees in psychiatry (rather than psychiatrists of many years' standing), and there is a trend for more experienced psychiatrists to achieve higher consistency. Only the British studies using the PSE have considered the impact of the experience of the clinician rater on the outcome in reliability studies. Other studies have used psychologists and one study involved psychiatric social workers as the clinician raters but these raters have been using structured interviews.

The PSE and SADS were developed for clinical groups but have been used in non-clinical populations. There has been little attempt to test validity, with Leckman's 'best estimate' method with the SADS being one acceptable example. The DIS was developed for use in general population samples, but the first reliability study was also done on a clinical group. Subsequent "field validity studies" were carried out but were, in reality, reliability studies devised in parallel with data collection.

The consistency of rates in longer term reliability studies are affected by (i) the interval from initial interview to follow up, (ii) the severity of the fall-off in test-retest reliability over periods of days to index episode and whether help was sought, weeks after the initial interview, (iii) the time period under review (with higher reliability for lifetime diagnosis than shorter-term prevalence), (iv) the design of the study, whether using clinician versus lay interviewer or clinician versus clinician comparisons, and (v) whether comparing DIS to clinical interview or another structured instrument. Recent studies have

questioned whether the DIS alone can provide sufficient context for recall of depressive episodes.

There is still no agreement as to whether lay interviewers perform the same task as psychiatrists when undertaking a diagnostic interview. The DIS is designed to keep interviewer decision making to a minimum to circumvent this difficulty. It is probably slightly more reliable in the hands of a clinician, but this is not sufficient to cause concern or outweigh the advantages of the instrument, including its use by lay interviewers.

#### Selection of a case-finding instrument in the reported study

The study to be described in Chapter 6 aims to investigate rates of depression over time in a non-clinical and socially homogeneous group of young adults, (i) to establish prevalence of depressive experience; (ii) to examine changes in prevalence over time, with particular attention to sex differences and (iii) to isolate possible psychosocial risk factors.

The study design called for the use of a case-finding instrument to generate categories of depressive disorder but also allowed for other variables (e.g. trait depression scales or patterns of help-seeking behaviour) that could generate alternative definitions of caseness where appropriate.

As estimates of lifetime prevalence were required, the PSE was inappropriate.

The DIS was selected for a number of reasons.

(i) It was designed for use in general population studies, and the data could be compared to other studies using the instrument for estimating prevalence and providing categories for risk factor research.

(ii) It was easy to administer and readily comprehensible.

(iii) It was the only instrument with the potential for use by a lay interviewer at a later date.

(iv) The highly structured nature of the instrument allowed for modification for use by mail.

(v) Both DSM-III and RDC diagnoses could be generated. The DSM-III diagnoses of major depression and dysthymia are in line with categories used in general population studies reported in Chapter 3. RDC diagnostic categories (which include minor depression) allowed for a wider range of depression categories. While the RDC minor depression category was not originally included in the DIS-generated diagnoses, it is readily derived from questions in the DIS by using the two weeks' minimum duration and a symptom cut-off of 2 from the possible 8 symptoms (Chapter 6).

(vi) Training and access to other researchers with experience of the instrument was available in Sydney.

(vii) Data on its reliability were supportive at that time (i.e. in 1983).

### Note

While much of the material present is very critical of the DIS, it should be noted that many of the concerns illustrated in this chapter were published subsequent to its use in this study. Robins (1989) and Burke (1986) have reported the instrument to be adequate for its designated task, particularly when complemented with other data to increase reliable recall of depressive episodes.

### Definition of terms

This chapter will consider epidemiological aspects of sex differences in rates of depression. Epidemiology has been defined (Charney & Weissman, 1988) as "the study of distribution and determinants of disease in human populations". Such studies can generate information about rates (incidence and prevalence), variation in rates by person, time or place, and identification of risk factors. Relevant terms are defined by these authors.

The 'rate' may be defined as the number of persons affected with a disease, disorder or characteristic, per unit of population, per unit of time. Definition of a 'case' was discussed in Chapter 1 with the conclusion that caseness be defined by the researcher in relation to the characteristic or disorder under investigation (Copeland, 1981).

'Incidence' refers to the number of new cases of a disorder with onset during a defined time period, generally a twelve-month period. 'Incidence rate' or 'new case rate'

refers to the number of new cases of a disorder over a period of time as a proportion of the population at risk for developing the disorder. 'Prevalence' refers to the number of cases, both old and new, found in a population in a defined time period. 'Point prevalence' defines the proportion of a population with the disorder under investigation at a given point in time; 'period prevalence' defines the proportion of a population with the disorder over a defined period (e.g. one month, one year) and 'lifetime prevalence' defines the proportion of the population alive at that time that has ever had the disorder under investigation. 'Prevalence rate' refers to the total number of cases of a disorder at a given time as a proportion of the total population.

A 'risk factor' is a specific characteristic or condition that seemingly increases the probability of present or future occurrence of a specific disorder. Such factors may be defined in terms of time, place or person. Time factors include definition of the lifetime risk (the proportion of a population expected to develop a disorder or characteristic, in past, present and future, including those who are deceased). 'Morbidity risk' is defined as an individual's lifetime risk of having a first episode of disorder. Definition by place may be used if an investigator isolates a particular population or region. Personal risk factors include socio-demographic variables, biological and physiological variables and personal habits.

### Review of sex differences in depressive experience

From the historical overview in Chapter 1, it is evident that the concepts concerning categorisation of depression have undergone considerable change over time but at an accelerated pace since World War II. The two World Wars had an impact on psychiatric theory in several important ways (Weissman et al, 1986). Firstly, there was an opportunity to witness the emotional performance of large numbers of healthy, young people under conditions of great stress, demonstrating the variety of normal

responses. Secondly, the upheaval caused by the wars led to changes in the economic and educational opportunities available to women and a subsequent period of questioning of sex roles and sex differences. Thirdly, the movement of European psychiatrists to USA led to cross-fertilisation of concepts and increased awareness of differences, leading to the trans-Atlantic studies mentioned in Chapter 1.

Weissman and Klerman (1977) discussed changes in women's roles in their introduction to a comprehensive review which attempted a broad coverage of English language studies reporting rates of depression in both inpatient and outpatient facilities, community surveys, studies of attempted and successful suicides, and studies of grief and bereavement, with the particular aim of reviewing the epidemiology of sex differences. They noted the diversity of use of the term 'depression', the lack of clarity of boundaries between the different concepts and problems of case definition. During the 1960s and 1970s, rates of depression were determined either by diagnosis in the clinical setting, or by counting hospital in-patient numbers. Both these study designs have the limitation of ignoring those people who had not presented for psychiatric treatment. Further, the use of such self-report measures as the Self-Rating Depression Scale or SDS (Zung, 1965) for case-finding does not allow generation of diagnostic categories nor the determination of the number of depressive episodes of clinical severity. The British had developed the Present State Examination or PSE (Wing et al, 1974) discussed in Chapter 1 as a standardised case-finding instrument, but the Americans were not using a structured case-finding instrument at this stage.

After review of the available literature, Weissman and Klerman noted an overall female to male sex ratio of about 2:1 in depressive experience, which they judged was consistent despite difficulties in case definition. However, the sex ratios were not consistent for all diagnostic categories, being almost equal (1.2:1) for rates of manic-

depression (bipolar disorder) and only in the order of 2:1 for non-bipolar depression. As non-bipolar depression occurs far more frequently, the sex ratios in this category tend to dominate. They noted no differences in grief and bereavement, which they took to indicate that grief was "qualitatively different from clinical depression".

### Sex differences in rates of bipolar disorder

By 1981, the category "primary depressive disorder" had been divided into bipolar disorder (which was seen as a discrete category), and other depressive disorders (which were combined into a heterogeneous category termed 'non-bipolar'). Boyd and Weissman (1981) reviewed the prevalence of depressive disorders in this context and found that the morbid risk (for industrialised countries) for bipolar disorder was of the order of 0.24% to 0.88%. These rates have been largely unaltered by data from subsequent studies (Robins et al, 1984; Bland et al, 1988) although the Epidemiological Catchment Area (ECA) study quotes rates as high as 1.2% (Weissman et al, 1988). In the same review, incidence rates ranged from 9.2 to 15.2 new cases per 100,000 per year for men and 7.4 to 32.5 new cases per 100,000 per year for women. When rates for first hospitalisation per 100,000 per year in three studies (in Denmark and U.K.) were compared, there were remarkably similar rates (3.9 to 10.7 for females, 3.0 to 8.3 for males).

Despite some suggestions that there are sex differences in both unipolar and bipolar disorder associated with genetic abnormalities linked to female chromosomes (Gershon and Bunney, 1976; Winokur and Tanna, 1969), the general consensus is that the female to male/sex ratio for rates of bipolar disorder is in the range 1:1.2 to 1.3:1 with non-significant sex differences (Boyd and Weissman, 1981; Robins et al, 1984; Bland et al, 1988) and, since 1981, interest in sex differences has focused on unipolar depression.

### Sex differences in rates of unipolar depression

Boyd and Weissman (1981) also considered studies of non-bipolar or unipolar depression to that time. They considered 16 studies and found a **point prevalence** of 1.8 - 3.2% for males and 2.0 - 9.3% for females in industrialised countries, and a higher rate (14.3% for males and 22.6% for females) in an African village (Orley & Wing, 1979) but the significance of this finding is unclear as the study results were somewhat at variance with the others noted. One study had used SADS/RDC (Weissman and Myers, 1978), a further three had used PSE (Henderson et al, 1979; Wing, 1977; Orley & Wing, 1979), and the rest had relied on clinical interview. The rates for those studies using a case-finding instrument are reported in Tables 3.1 and 3.2.

For **life-time prevalence** rates, there were only two studies using SADS/RDC criteria, the other four studies coming from Scandinavia. All included treated and untreated cases and the authors noted "remarkably consistent" rates ranging from 8%-12% for men and 10%-26% for women in the studies using RDC and a much wider variation (2%-12% for men and 5%-26% for women) when clinical diagnosis alone was used. They cited annual incidence rates based on a review of nine studies examining case registries of people presenting for treatment of depression for the first time. Rates vary from 130-201/100,000 for men (with rates of 27-65/100,000 for psychotic depression) and 330-500/100,000 for women (37-123/100,000 for psychotic depression). However, these figures are likely to be an under-estimation of rates for men, as women are known to present for treatment more often than men. In a "representative national sample" of 2264 adults, 31% of females and 23% of males reported seeking help from mental health agencies, but college educated males were more likely to seek help than less well educated males and at the same rate as females (Veroff, 1981).



Table 3.1      One month prevalence rates per 100 subjects  
for depressive episodes using PSE

Reference	n	Males %	Females %	Total %
Camberwell, U.K. Wing et al, 1977	800	4.8	9.0	7.0
Canberra, Australia S.Henderson et al, 1979	756	2.6	6.7	4.8
Outer Hebrides, U.K. Brown et al, 1981	169	2.6	4.5	4.5
Uganda, Africa Orley and Wing, 1979	206	14.3	22.6	-
Cantabria, Spain Vazquez-Barquero et al, 1987	1223	4.4	7.8	6.1
Nijmegen, Holland Hodiamont et al, 1987	3232	-	-	5.4*

\* Rate for all PSE cases was 7.2 for males, 7.5 for females,  
7.4 total

Table 3.2 Prevalence rates/100 subjects for major depression based on community surveys using RDC and DSM-III diagnostic criteria

Place and time of study, and investigator(s)	n	Age range (yrs)	Diagnosis	Males %	Females %	Total %
<u>Current prevalence</u>						
Stirling County, Canada 1952, Murphy (1980)	2125	18+	DSM-III*	-	-	4.1
North Carolina, USA, 1972 Blazer & Williams (1980)	997	65+	DSM-III*	3.2	4.0	3.7
New Haven, USA, 1975 Weissman & Myers (1978)	511	26+	SADS/RDC	3.2	5.2	4.3
Edinburgh, 1983 Dean et al (1983)	576	18+	SADS/RDC	-	7.0	-
Zurich, 1984 Angst & Dobler-Mikola (1984a)	591	23-24	DSM-III*	0.4	3.1	1.8
* DSM-III diagnoses derived from original data						

Verbrugge (1985) also reported that where psychosocial factors were allowed to operate (i.e. for chronic, less severe illnesses) that sex differences in help-seeking behaviour were greatest. There are sex differences in choice of agency in that that men are more likely to turn to specialist mental health services and women to general practitioners (Shapiro et al, 1984; Temkin-Greener and Clark, 1988).

The evolution of case-finding instruments has led to improved reliability of case-finding rates, but there has been some difficulty in comparing prevalence rates as the various case-finding instruments have differing time frames for prevalence data.

#### Review of general population studies using the PSE

Several general population studies have used the PSE (which reports one-month prevalence rates) and Table 3.1 shows the rates for depression in these studies.

Following Henderson's example in Canberra (Henderson et al, 1981), other studies (Hodiamont et al, 1987; Vazquez-Barquero et al, 1987) have used a two-stage screening procedure, first giving the GHQ as a screening instrument and then the PSE for those subjects who score at or above threshold for caseness. As women are known to produce higher scores on self-report instruments, there are problems in using such a instrument as a screening device as one would expect a bias favouring female intake after the screening stage. Henderson's study in Canberra did not establish significant differences by age or sex, reporting GHQ scores that generated a "probability of caseness" of 8.2% for males and 8.8% for females and an overall point prevalence (using PSE) of 9.0% +/- 3.2% for threshold and definite cases with no significant sex differences. Rates for definite cases were 1.1% for males and 1.8% for females, but a trend towards female preponderance was shown when 'threshold cases' are included



Table 3.4 Prevalence/100 subjects for major depression based on Community Surveys using RDC and DSM-III diagnostic criteria

Place and time of study, and investigator(s)	n	Age range (yrs)	Diagnosis	Males %	Females %	Total %
<u>Lifetime prevalence</u>						
New Haven, U.S.A., Weissman & Myers (1978)	511	26+	SADS/RDC	12.3	25.8	18.0
National Survey, Deviant Behaviour, 1983 Elliott et al (1985)	1496	18-24	DSM-III /DIS	8.4	-	8.4
Stirling County, 1952 Murphy (1980)	1003	18+	DMS-III*	-	-	16.0
<u>E.C.A. Study, U.S.A., Robins et al (1984)</u>						
New Haven	5034	18+	DSM-III	4.4	8.7	6.7
Baltimore	3481	18+	DSM-III	2.3	4.9	3.7
St. Louis	3004	18+	DSM-III	2.5	8.1	5.5
Edmonton, Canada, 1988 Bland et al (1988)	3258	18+	DSM-III	5.9	11.4	8.6
Christchurch, NZ Wells et al (1989)	1498	18+	DSM-III	3.8	9.0	6.4

\* Reanalysis of original data to meet DSM-III criteria

Table 3.5 Prevalence/100 subjects for dysthymia based on Community Surveys using DSM-III diagnostic criteria

Place and time of study, and investigator(s)	n	Age range (yrs)	Males %	Females %
<u>Lifetime prevalence</u>				
<u>E.C.A. study, USA Weissman et al, 1988</u>				
New Haven	5011	18-44	3.2	4.3
Baltimore	3333	18-44	1.0	3.2
St. Louis	3970	18-44	2.6	5.2
Piedmont	3825	18-44	1.3	2.9
Los Angeles	3109	18-44	3.6	4.8
Edmonton, Canada Bland et al, 1988	3258	18+	5.2	2.2*
Christchurch, NZ Wells et al, 1989	1498	18+	3.8	9.0**

\* Total rate for 25-44 age group = 3.8/100

\*\* Total rate for 25-44 age group = 5.5/100

giving rates of 5.9% for males and 9.2% for females (Henderson et al, 1979). Marital state was more indicative of caseness than age or sex, with higher rates noted in the widowed, single and separated subjects.

A study of a Dutch urban area (Hodiamont et al, 1987) collected a random sample of 3,232 subjects who completed the GHQ, of whom 2,486 (with a GHQ score of 10/30 or above) were interviewed, using the PSE. There were no significant sex differences in rates of PSE caseness, being 7.2 +/- 2.0% for men and 7.5 +/- 2.0% for women which corresponded with Henderson's finding, but not with Wing's study in Camberwell (Bebbington et al, 1981) which found rates of 6.1% for men and 14.9% for women, with a significant sex difference. Another study in a rural community in Cantabria, Spain (Vazquez-Barquero et al, 1987), found rates similar to Wing's study, with 8.1% of men and 20.6% of women being cases. In this study, the men had slightly higher rates of endogenous depression (2.6% for males; 2.1% for females) and the women had higher rates of neurotic depression (1.7% for males; 5.7% for females) and anxiety neurosis (2.4% for males; 6.1% for females). The rates were inversely related to socioeconomic status for both sexes. There was no significant association between psychiatric caseness and marital state but presence of 3 children under 14 was correlated with psychiatric caseness in both sexes. Orley and Wing's (1979) study using the PSE in two Ugandan villages, also found high rates of caseness (20.4% of the population) with depressive disorders in 14.3% of men and 22.6% of women. All these studies other than the Canberra study report a two-fold excess of depressive experience in women.

Data from the ECA study reporting one-month prevalence of depressive disorders (Regier et al, 1988) can be compared to the studies using the PSE. For

combined DIS disorders, women (16.6%) had higher rates than men (14.0%), which reached statistical significance in all age groups except the 18-24 year age group. For affective disorders, the rates for women rose from 5.3% in the 18-24 age range to the highest rate of 8.2% in the 25-44 year age range, while male rates in the 18-24 year age group were 4.5% and did not change significantly over the same time period. The authors described a bell-shaped curve, with a peak for depressive experience in the 25-44 age range and a similar trend occurring with anxiety disorders in women. Much of the female excess was accounted for by the significantly higher rates for dysthymia in females, which decreased after the age of 64, when sex differences in affective disorder also diminished (Weissman et al, 1988a; 1988b).

#### General population studies using the SADS/RDC system

This measure was used in the Mental Health-Clinical Research Branch Collaborative Program on the Psychobiology of Depression Clinical Study (Katz et al, 1979) in which 523 probands and 2,289 relatives were studied with aims of refining research nosology, investigating neurobiological functions operating in depression and defining areas for future research. It was not a general population study as such, although the extensive interviewing of relatives produced a non-clinical sample. Data from this group were reported in a birth cohort study in which younger cohorts (in the 24-44 age group at the time of the study) reported higher lifetime rates of major depression than cohorts of older subjects (Klerman et al, 1985).

An epidemiological study of 511 people in New Haven, USA (Weissman and Myers, 1978) was designed to derive point and lifetime prevalence of affective disorders using the RDC. Point prevalence rates for probable and definite major depression (4.3%) and minor depression (2.5%), with a female:male sex ratio of 1.6:1 for major depression and 1.2:1 for minor depression were reported. Lifetime rates

(probable and definite) were 20% for major depression and 9.2% for minor depression, with a lifetime rate of 26.7% for major or minor depression, reducing to 24.7% if only 'definite' categories were included. They also reported grief reactions in 2.7% of men and 16.2% of women.

The New Haven figures should be viewed with suspicion as the threshold for RDC 'probable' minor depression (which requires a minimum of one week's duration plus at least two symptoms) is so low that it seems counter-intuitive that the lifetime rates for minor depression should be much less than those quoted for major depression. The likely explanation is that the depressive episodes with short duration and fewer symptoms are seen as part of a normal human experience and not necessarily remembered as pathological, or are without sufficient impact on the subject's experience to be reliably recalled. The large sex difference in rates for grieving also seems unusual, being contrary to trends in other studies reviewed by Nolen-Hoeksema (1987) where no sex differences in depression associated with grieving were noted.

Weissman (Weissman, Leaf, et al, 1988) later stated that the SADS/RDC study was seen as a pilot to demonstrate the utility of a standardised case-finding instrument and several factors (need for clinician raters, lack of ability to derive DSM-III diagnoses and lack of computerised scoring for the instrument) determined that interest turned away from RDC/SADS system towards the implementation of the DIS, which overcame these problems. However, the multi-centre collaboration involved in the Clinical Studies Program laid the groundwork for the ECA study.

#### General population studies using the DIS

Several general population studies have used the DIS as a case-finding instrument. Tables 3.2 (current prevalence), 3.3 (six and twelve month prevalence), and



3.4 (lifetime prevalence) summarise rates for major depression in general population studies and Table 3.5 summarises data for lifetime prevalence of dysthymia from some of the same studies for males and females, with particular reference to young adults.

The ECA study has already been noted. In a paper (Weissman et al, 1988) reporting multiple prevalence rates for affective disorders from the five ECA study sites, they noted that two-week prevalence rates for major depression ranged from 1.0% to 1.8% and lifetime prevalence rates ranged from 2.9% to 4.4%, all of which are lower than in previous studies reviewed. These rates require further investigation as approximately half the cases reported were still current or had occurred within the previous twelve months suggesting that earlier episodes may have been forgotten. This study did not find any significant sex differences in bipolar disorder at any of the five sites but found a higher female:male sex ratio for major depression at every site in every age group. The size of the sex difference varied with age, being highest in the 18 to 44 year age group. Lifetime prevalence rates for dysthymia varied from 2.1% to 4.2%, with a similar pattern of sex ratio findings as those for major depression (for dysthymia, the female:male ratios varied from 1.5:1 to 3:1 across centres and sex differences were greatest in the 45-64 age group). When all DIS diagnoses are combined (including antisocial personality and alcohol abuse/dependence), the lifetime prevalence rates for any disorder for males ranged from 30.6% to 39.6% and for females, from 25.7% to 36.7%, with a non-significant trend towards a male preponderance.

A study (Bland, 1988) in Alberta, Canada, used the DIS in a sample of 3,258 community residents. For major depression, they found lifetime rates of 5.9% for men and 11.4% for women and, for dysthymia, rates of 2.2% for men and 5.2% for women with significant sex differences in both categories. There was a female:male sex ratio of 1.9:1 for affective disorder and 1.7:1 for anxiety/somatoform disorder. They

reported lifetime prevalence rates for all DIS-generated diagnostic categories (which included drug and alcohol dependence) of 40.7% for males and 26.8% for females, with a significant male preponderance. When alcohol/substance abuse and simple phobia categories were removed, the rates for all remaining disorders were 19.1% for males and 23.6% for females, now with a significant sex difference in the reverse direction.

A further study of 1498 randomly selected adults in Christchurch, New Zealand, reported six-month prevalence rates (Oakley-Brown et al, 1989) for major depression of 5.3% (3.4% for men and 7.1% for women), and lifetime rates (Wells et al, 1989) for major depression of 12.6% (8.9% for men and 16.3% for women); and for dysthymia, 6.4% (3.8% for men and 9.0% for women). Although there were significant sex differences for all affective disorders, after Bonferroni correction a significant sex difference was demonstrated only for dysthymia. No other study has commented on the use of this statistical technique. These authors also note that males show a statistically significant greater lifetime prevalence rate when all DIS categories are combined (39.6% for males, 33.7% for females). They demonstrate that sex differences can be found in either direction depending on which diagnostic categories are compared. If all categories other than drug and alcohol abuse are combined, the lifetime prevalence rates are 15.7% for males and 29.9% for females ( $p < .001$ ). If all disorders other than phobias are combined, the rates are 38.7% for males and 27.3% for females ( $p < .001$ ).

Both the Canadian and New Zealand studies have reported that for categories of major depression, dysthymia, anxiety disorders and alcohol and substance abuse, the highest prevalence rates (both for the preceding six-month period and lifetime) occur in subjects aged 18 to 44 years and decrease for subjects over 65 years of age.

In another study, the DIS was administered to a random sample of 1551 community residents in Puerto Rico (Canino et al, 1987a). Here lifetime prevalence rates for major depression were 3.9% for men and 5.8% for women and for dysthymia, 1.9% for men and 7.5% for women, with statistically significant sex differences (Canino et al, 1987b).

### Review of mental health studies using a longitudinal design

The Lundby study that commenced in Sweden, in 1947 (Essen-Moller, 1956), has already been described in Chapter 1. Further field examinations of this group took place in 1957 and 1972 (Hagnell et al, 1982). Depression was described as mild, medium or severe, with mild depression being defined by the presence of depressive symptoms and a 50% reduction in normal activity such that it would entitle the subject to sickness benefit at the time of the episode. Over the 25-year time period this study indicated a statistically significant increase in the incidence of depression of medium and severe levels of impairment. Both incidence and morbid risk were calculated for subjects who were grouped in cohorts by age. They reported incidence for severe depression of 0.4/100,000/year (men) and 0.9/100,000 (women) in the 1947-1957 time period of 0.3/100,000/year (men) and 0.4/100,000/year (women) in the 1957-72 time period. Rates for probability of a first episode of any type of depression were 0.11% (men) and 0.30% (women) in the 1947-1957 time period and 0.26% (men) and 0.49% (women) in the 1957-72 time period. For both sexes, there was a statistically significant increase in morbid risk of depressive experience over time, the risk being greatest for men in the 20 to 39 year age group having an episode of medium or severe depression, and a trend towards a more equal sex ratio, both for incidence and probability of first episode. They note that "the increase of depressive disorders.....among men in their twenties and thirties is an alarming phenomenon."

A study of those presenting for psychiatric treatment in Finland (Helgason et al, 1977) reported a total incidence rate (all disorders) of 506/100,000/year for males and 607/100,000/year for females. A higher percentage of males in the cohort was hospitalised (19.8%) than of females (16.8%) during the seven-year course of the study. Incidence rates were highest for males in the 20 to 54-year age group, with a peak at the ages of 45 to 49 years. Females experienced the highest incidence rate from 20 to 44 years, with a peak at 30 to 34 years. Rates for non-bipolar depression were 110/100,000/year (55/100,000/year for males and 152/100,000/year for females).

Angst and his group in Switzerland conducted a four-year longitudinal study (Angst et al, 1984a) commencing in 1978, comprising 299 women and 292 men selected from a general population group (mentioned in Chapter 1). Angst's group used a structured interview, the SPIKE, to generate three- and twelve-month prevalence data in the first part of a longitudinal study. They quoted three-month prevalence data for DSM-III major depression (men 1.9%, women 3.4%), and for their own category, Extended Depressive Episode (EDE) (men 4.6%, women 4.3%) and twelve-month prevalence for major depression (men 2.3%, women 11.2%), and EDE (men 4.7%, women 10.7%). They concluded that the three-month figures were more likely to be reliable but show that sex differences in rates vary with the definition and time-frame used. To explain the differences in sex ratios between the 3-month and 12-month estimates, they postulated that men more readily forget episodes.

The Swiss group was interested in incidence and prevalence rates of depression and anxiety, sex differences and case definition. They developed two approaches (Angst et al, 1984c) to define depression, one using role impairment and the other using the number of symptoms present, with females requiring more symptoms for the diagnosis of depressive disorder than males.

The Stirling County study (Leighton et al, 1959) was also discussed in Chapter 1. Lifetime prevalence rates of anxiety and depression were measured in 1952 and 1970. Murphy re-analyzed the original data to report episodes of anxiety and depression in terms that could be related to DSM-III diagnostic categories (Murphy et al, 1985). While lifetime prevalence rates for depression remained essentially unchanged over the three decades of the study, rates for women had decreased from 17% to 15%, while rates for men had increased from 8% to 10% with a trend towards more equal sex ratio. The sex differences were least in the 40 to 69 year age range, where rates for women were around 16% and for men, around 14%.

The Midtown Study (Srole, 1962) similarly noted a trend for improvement in current prevalence of "mental morbidity" of men and women in the 40 to 44 year age group. From 1954 to 1974, the rates for males in cohorts remained stable at 9% while rates for females changed from 21% in 1954 to 8% in 1974 leading to an equal sex ratio in the later cohort (Srole et al, 1978).

Findings from longitudinal studies suggest that subjects born in the three most recent decades have a higher rate of depression and concur with research from other sources looking at the effects of recency of birth in different age cohorts. Klerman's group have noted similar findings when they reported the results of interviews of 2289 relatives of 523 probands as part of a large NIMH study (Klerman et al, 1985) in the U.S.A. where the group was categorised by decade of birth. Here they found a predominance of female depressives in all birth cohorts, but an increase in depressive experience in men, so that the male-female differences had both fluctuated over time and lessened in the last three decades.

Hasin and Link (1988) have drawn attention to some of the methodological problems associated with cohort effects, such as selective mortality (those with major depression dying younger and hence not being present in older cohorts), memory effects (with older subjects being less able to recall episodes of depression which occurred when they were younger) and younger people being more likely to give a psychological explanation for episodes. These factors are coupled with the possibility that older people may be more likely to attribute depression to physical causes or to have physical illnesses complicating their depressive episodes. They tested these propositions by asking a series of 152 people, randomly selected from the telephone book, to imagine themselves in a hypothetical situation simulating a depressive episode and to consider their mood state in such a situation. They found that older people were much less likely to perceive a depressive episode as a psychological or emotional problem than younger residents of the same area and concluded that the recognition of depression as an emotional or psychological problem influences the amount of recall of such episodes. This study also has implications for research into sex differences, for if women are more attuned to emotional problems they are also more likely to recall episodes of depression.

Jorm (1987) focused on the specific question of age of onset. After commenting that most previous reviews had only compared total sex ratios with little reference to age-specific data, he employed a technique of "quantitative integration of research data" to compare those studies using both self-report questionnaires and interview schedules to define depression, in which age ranges were also reported. He found that at the extremes of the age range, in children (with a prevalence of 8%), and the elderly (with a rate of 17% in the over-70 age group), there were no sex differences. In the 20's age group, he reported rates of 17% for males, then a steady rate between 11% and 12% from the age of 40; for females, this rate was highest (at 26%) in the 20's, falling to 22%

by 40 and 17% by the 70's. He concluded that the sex differences are age-specific, and most apparent for less severe forms of depression in women during the childbearing years. These findings have been reinforced by subsequent data from the DIS derived-studies where subjects have been categorised by age group (Regier et al, 1984; Bland et al, 1988; Wells et al, 1989).

Von Zerssen and Weyerer (1982) hypothesised that the age effect can be explained by the change in diagnostic labelling, so that disturbed behaviour, particularly in boys, is labelled depression in childhood (requiring psychiatric intervention) but labelled personality disorder or criminal behaviour (requiring legal intervention) from adolescence on and later. The implication is that differences in sex ratio are influenced not only by overall morbidity but also by definition of caseness. They also noted that the female preponderance in depression is not seen in developing countries or Japan, and has only been reported in Western countries since World War II. They found no answer to the question of whether this was due to changes in morbidity rates or case identification, but noted that earlier studies excluded minor disorders from case definition and that earlier investigators had relied on reported behaviour rather than subjective reports of distress, with the former approach favouring male preponderance. They studied 1536 subjects in Bavaria (688 males, 848 females) using a semi-structured interview (Goldberg et al, 1970) to generate ICD-8 diagnoses. They found significant sex differences for unipolar endogenous depression and bipolar depression (higher in females), and drug abuse (higher in males), which upheld their hypothesis that female psychopathology may be overestimated because of exclusion of aberrant social behaviour from case definition in adults.

Table 3.6      Point prevalence rates/100 subjects  
for depressive symptoms

Reference	Self-report assessment	Males %	Females %	Total %
Weissman & Myers, 1978 New Haven, U.S.A.	8-item Index	12	20	16
Weissman & Myers, 1978 New Haven, U.S.A.	8-item Index	16	20	18
Comstock & Helsing, 1976 Missouri, U.S.A.	CES-D	16	22	20
Comstock & Helsing, 1976 Maryland, U.S.A.	CES-D	12	21	17
Blumenthal, 1975 Pennsylvania, Indiana, Kentucky & Michigan, U.S.A.	Zung	12	32	-
Eaton & Kessler, 1981 Los Angeles, USA	CES-D	11	21	16
Warheit et al, 1973 Florida, USA	18-item Index	13	24	-



### Sex differences in rates of depressive symptoms

Prior to use of structured case-finding interviews, most psychiatric, psychological and sociological studies used self-report questionnaires (such as the Beck Depression Inventory or BDI, (Beck et al, 1961) and the Self-rating Depression Scale or SDS, (Zung, 1965)) to derive dimensional data rather than diagnostic categories. Higher rates of moderate to severe symptoms of depression are found when self-report questionnaires are used (rather than structured interviews) but the female:male ratio of 2:1 persists in most groups. This phenomenon can be seen by comparing data summarised in Tables 3.1 and 3.2 (current and one month prevalence rates for diagnostic categories) with Table 3.6 (point prevalence rates for depressive symptoms). The symptom rates are roughly four times those for categorical diagnoses for both sexes. As there is not a strong relationship between scoring highly on a self-report depression scale and meeting criteria for RDC or DSM-III depressive categories (Charney & Weissman 1988), the data from self-report questionnaires will be considered separately. Table 3.6 summarises point prevalence data from seven studies, reporting rates from 10% to 19% for males and 20% to 31% for females.

Sex differences on self-report measures diminish with age, as Ensel (1982) found in a study using a self-report scale, the Center for Epidemiological Studies Depression Scale or CES-D (Radloff, 1977), in subjects over 50 years of age. Radloff (1985), using the same measure found that there were no sex differences in mean scores for the under-40 age group, with highest scores for both sexes in the 18-24 year groups. Males had higher mean scores than females in the over-65 age group (Radloff, 1980). Zung (1967) had earlier suggested use of a higher base-line level in self-report symptom scales in elderly subjects to allow for an overall decrease of subjective feelings of optimism and increase of psychosomatic complaints. In older subjects, the increased likelihood of physical illness (with such symptoms as tiredness, appetite disturbance

'and concentration difficulties) can also have a confounding effect on rates of depressive symptoms.

#### Data from in Australian studies reporting rates of depression

Berah (1983) reviewed Australian studies to determine whether the same sex differences for depression reported in the world literature were apparent locally. The data sets were derived from statistics of patients admitted to the state psychiatric system and general hospital psychiatric units until 1970 (which were compared with a three-month period in 1981), two community surveys carried out in the rural town of Heyfield (Krupinski et al, 1970) and two Melbourne suburbs (Krupinski and Stoller, 1971; Krupinski and Mackenzie, 1979), and a nationwide general practice morbidity survey (NH & MRC, 1962-3). These data showed that, prior to 1978, there were more male in-patients, although the male:female ratio rose 1.1:1 in 1969 to 1.4:1 in 1978. Over the same period, outpatients' figures showed a male:female ratio of 1:1.2 in 1969 changing to a ratio of 1.6:1 in 1978. While the numbers were small (n=97 in 1969, n=137 in 1978), this may indicate an increasing acceptance of psychiatric services by males or an increase in psychiatric disorder of mild to moderate severity in males. In general hospitals, there was a consistent female preponderance in admission rates (male:female ratio of 1:1.3 in 1962 and 1:1.8 in 1981). The General Practitioner Morbidity Survey revealed equality in sex ratio of general practice attenders (52% of all illness episodes reported by females) but a female preponderance (63%) in episodes of psychiatric illness. There were consistently higher rates of ICD-9 depressive neurosis across all treatment facilities.

The combined data from all the surveys (using ICD-9 diagnostic categories) reflected the trends noted worldwide, with males being more frequently diagnosed as having alcoholism or personality disorder and females more frequently diagnosed with

neurotic depression or another neurosis. There were no sex differences in rates of functional psychoses.

Berah concluded that there were more women with both treated and untreated psychiatric morbidity, excluding alcoholism, and stated that the sex ratios vary depending "on how the boundaries of the category are drawn". This conclusion is in keeping with rates for psychiatric morbidity already discussed in relation to overseas studies.

Australian suicide rates for the periods 1969-73 and 1976-80 were compared (Dorsch and Roder, 1983). Rates for males were generally at least twice those for females, with higher rates in the middle-aged and older groups. In considering the 20-29 year age group, rates for females fluctuated over the two time periods but remained essentially constant while rates for males were 24% higher at the end of the study. The authors considered that rising unemployment rates may have been implicated in the increase in suicide rate in the young.

The N.S.W. Health Commission (Reynolds et al, 1979) undertook a study to examine the prevalence of psychosocial problems and to consider implications for intervention, using 37,678 adults presenting to a health screening facility from 1976 to 1977, in which a random sample of 512 were interviewed more extensively. All subjects were asked whether they had periods of sadness or depression, to which 5% of females and 2% of males stated they were depressed "often and for long periods" (yielding 54 males and 85 females) and 30% of females and 15% of males were depressed "often and for brief periods". The prevalence decreased with age but was consistently greater for females. Those who had "often and long" episodes were asked whether these were a problem, to which 11% of both sexes stated the episodes were "a

great problem". When asked about causes of depression, the most frequent responses in men were lack of purpose (26%), sense of failure and lack of achievement (26%), divorce and separation (26%), loneliness (17%), work (13%), marriage/home situation (9%). Women rated marriage/home situation more highly (30%) than loneliness (24%), lacking purpose in life (15%) and bereavement (15%). About 40% of both males and females had discussed their depression with a health professional, generally a family doctor. Seventeen per cent of males and 12% of females had seen a psychiatrist and of these, 83% of males and 88% of females had been prescribed anti-depressants and currently, 4% of males and 6% of females were taking them regularly. The study did not separate clinically relevant depression from understandable sadness but the levels of help-seeking and ingestion of anti-depressants would indicate that these subjects would generally be seen as 'depressive cases'. There is a further problem in that predictor and outcome variables were confounded (e.g. 'lacking purpose in life' is as likely to be a consequence of depression as a cause) but it does illustrate some reported sex differences in the Australian context.

#### Groups in which there have been a consistent lack of sex differences in depression

The Midtown Manhattan study (Srole, 1962) reported symptoms on a continuum from health to ill-health rather than diagnostic categories. They reported the distribution thus: 18.5% well; 36.3% mild symptom formation; 21.8% moderate symptom formation; 7.5% marked symptom formation; 13.2% severe symptom formation; 2.7% incapacitated. This meant that just over half (54.8%) their population were well-functioning and 23.4% were judged as impaired and 21.8% as borderline, with moderate symptom formation. They reported no significant sex differences in these categories but noted an increase in symptomatology with age and, more significantly, with lower socio-economic class.

Studies of college and university students have failed to demonstrate sex differences in rates of depression. Hammen and Padesky (1977) and Parker (1979) reported no sex differences on administration of self-report questionnaires to groups of university undergraduates but these findings were downplayed as the groups were deemed to be too socially homogeneous (Weissman and Klerman, 1977; Nolen-Hoeksema, 1987). However there have since been further studies (Bryson & Pilon, 1984; King & Buckwald, 1982) that have confirmed those earlier reports. The mean age of the student groups tends to lie in the early 20's, prior to the 25 to 40 year range in which women are at their most vulnerable (Jorm, 1987).

In a study using both self-report questionnaires and interview schedules for diagnosis of depression, a group of 298 volunteers (117 male, 181 female) selected by random telephone dialling used the Diagnostic Interview Schedule (DIS) and the Beck Depression Inventory (BDI) to measure rates of depression (Oliver and Simmons, 1985). The point prevalence rates for DIS-generated DSM-III major depression were 5.2% for males and 6.6% for females without significant sex difference, while for lifetime prevalence rates were 12.8% for males and 23.8% females, with a significant sex difference. As in other studies, the females tended to score more highly on the self-report measure (BDI). The authors suggested that the difference in lifetime prevalence rates was explained by females having a greater tendency to have further episodes of depression, while men were more likely to have a single episode. However, these findings could also suggest that men fail to report episodes of depression; or fit Angst's hypothesis that men are more likely to forget episodes of depression over time.

Jenkins (1985) tested the hypothesis that there were no sex differences in minor psychiatric morbidity, in symptom profile or outcome over twelve months in a "population of men and women who were as comparable as possible in every

demographic, social and occupational respect". She selected a group of 321 young civil servants (138 women, 183 men) and used the GHQ as a screening instrument (with a score of four or more as a cut-off for caseness), then the Clinical Interview Schedule or CIS (Goldberg et al, 1970) to interview the 184 identified as probable cases. Overall, there were no sex differences in rates for caseness, (with a prevalence of 34.3% in women and 36.3% in men) but men showed a better recovery rate at the twelve-month follow-up though this comparison failed to reach statistical significance. The male/female sex ratio was 0.93 using GHQ scores and 0.94 using CIS scale scores, indicating a slight, though non-significant trend for women to score more highly than men on each scale. There were significant sex differences in self-perception of illness and illness behaviour, with women reporting more depression, somatic symptoms, tiredness and irritability, and men more problems with concentration, retardation and anxiety. Only the differences in somatic symptoms and retardation reached statistical significance. Jenkins concluded that "constitutional factors do not play an important role in the excess of minor psychiatric morbidity reported in women, and that environmental strains are more likely to be of importance". She did question whether apparently healthy women may experience more fluctuations in mood and associated somatic symptoms than apparently healthy men.

A Finnish study (Holmström et al, 1987) of 200 young men and women (47% female, 53% male) commenced at the beginning of their university studies (in 1965) and followed them through the start of their working life (from 1978 to 1979). They used the Minnesota Multi Personality Inventory (MMPI) to measure such variables as ego strength, anger and self-esteem, and a semi-structured interview to cover areas of work satisfaction, health, social interests and intimate relationships. The investigators had a particular interest in sex differences in social and personal adaptation. They found no sex differences in terms of "traditional psychiatric assessment" nor

"psychological quality of life" but the women reported more symptoms and tended to talk more easily about their problems, while the men reported fewer symptoms and were more likely to use alcohol as a coping device. While there were no sex differences in terms of academic performance, the women tended to have lower self-esteem levels during the transition phase from school to university and after starting their careers. The sex differences in terms of self-esteem in their post-graduate years seemed to be related to social status and power, and disappeared when the social status variables were controlled.

Egeland and Hostetter (1983) selected an Amish group for a study of the genetics of bipolar disorder and used the SADS-L to generate RDC diagnostic categories. They reported a slight male preponderance for bipolar disorder and equal sex ratios for major depression, and noted an under-representation of personality disorder, anxiety and minor depression. They considered that the lack of alcoholism and sociopathy amongst the men might constitute an explanation for the equal sex ratio and considered - but rejected - the proposition that the most aberrant males had left the community, thus artificially lowering the rates of personality disorders. They also considered the explanation later favoured by Nolen-Hoeksema (1987), that there were undiagnosed "silent depressives", women who were still managing to function and thereby avoiding being categorised. They stated that this may indeed be the case for minor forms of psychiatric disorder but that the nature of the Amish way of life for men and women was such that serious psychiatric disorder would be recognised because of the demanding work expected and the closeness of the community.

### Summary

There is widespread but not universal reporting of sex differences for unipolar depression, generally with a male: female ratio of around 1:2. Sex differences are not established for bipolar depression.

There are consistently higher prevalence and incidence rates for females if only depression and anxiety disorders are measured. There are consistently higher rates for alcohol/drug abuse and anti-social personality disorders in males. If alcohol/drug abuse and anti-social personality disorder are included in the rates for psychiatric disorders, the sex differences disappear. If substance abuse and personality disorders are excluded, there is a female excess in rates of combined psychiatric disorders, with the most significant sex differences in depressive and anxiety disorders. These findings suggest that overall rates of psychiatric morbidity may be the same for both sexes, but with different modes of expression in each sex.

The prevalence and incidence rates vary with the current age of the cohort. A number of researchers have suggested that there is an increase in rates of depression in younger age groups but there are some methodological problems that have been noted. Female preponderance is greatest for adults aged 20-40 years, although the magnitude of sex differences has been reported as diminishing in younger age cohorts.

The magnitude of sex differences is partly determined by the case-finding instrument used. The use of self-report measures can generate greater differences than operationalised diagnostic systems (like DSM-III) using structured interview schedules (like the DIS).



The magnitude of sex differences depends to some extent on the homogeneity of the sample in terms of psychosocial variables. Researchers investigating groups selected for their demographic and social homogeneity report an absence of sex differences in depressive symptom levels. This finding indicates that the differences, when present, may well be determined by psychosocial risk factors - not only for symptom levels but for defined 'cases' - the focus of this study.

## CHAPTER FOUR

### REVIEW OF RISK FACTORS TO DEPRESSION WITH PARTICULAR REFERENCE TO SEX DIFFERENCES

#### CONTENTS

Examination of 'artifactual' factors  
 Design of instruments and possible sex biases in use of  
 diagnostic criteria  
 Reporting of distress and depression  
 Help-seeking patterns  
 Examination of 'real' factors  
 Genetic factors  
 Biological factors  
 Psychosocial factors  
 Conclusions

#### Introduction

Possible explanations for the female preponderance of depressive experience can be divided into artifactual factors and real factors.

The artifact hypothesis proposes "that women perceive, acknowledge, report and seek help for stress and symptoms differently than men and that these factors account for the sex ratio findings" (Weissman & Klerman, 1977).

#### Examination of 'artifactual' factors

The sorts of artifactual factors that could account for sex differences in depression include sex biases in design of instruments and diagnostic criteria; differential patterns of help-seeking and reporting distress and depression that do not necessarily reflect the true prevalence in the general population.

#### **Design of instruments and possible sex biases in use of diagnostic criteria**

Weissman and Klerman raised the question of different stereotypes for normality in males and females but neglected to consider whether the authors of questionnaires had addressed sex role issues when designing their depression measures. This is important as it

was noted in the last chapter that sex differences are generally greatest when self-report measures are used and are obviously influenced by the wording of questions or by inclusion of sex-dimorphic items such as crying.

There is also evidence of bias in the diagnosis of personality disorders, even with the use of operational criteria. This is most apparent in the case of histrionic (described more commonly in females) and sociopathic (more commonly in males) personality disorder diagnoses and is found even when clinicians are presented with identical case histories for each sex (Warner 1975; Ford and Widiger, 1989). It is not clear whether any similar sex bias operates in the diagnosis of depressive symptoms.

The authors of DSM-III have tried to eliminate sex-biased references and included representatives with a feminist viewpoint on advisory committees. The feminist viewpoint of the time held that the arbitrary definitions of psychiatric disorder fitted with "masculine biased assumptions about what behaviour are healthy" and that any diagnosis occurring much more frequently in one sex should be viewed with great "suspicion and caution" (Kaplan, 1983). Spitzer replied that "we don't accept the basic principle that concentration in one sex equals bias" (Williams and Spitzer, 1983, Holden, 1986). Another study (Loring & Powell, 1989) found an interaction between race and gender in making diagnoses. Four psychiatrists (white male, black male, white female, black female) were given case summaries where the only details that changed were the sex and race of the patient. They found that patients given a diagnosis of dependent personality disorder tended to coincide with the sex and gender of the diagnosing psychiatrist, and that male psychiatrists were most likely to give a diagnosis of histrionic personality disorder to white females and paranoid personality disorder to black females. The male psychiatrists were more likely to assign females to depressive disorders and it was suggested that male clinicians may over-estimate the prevalence of depressive disorders in women. They also postulated that the literature on the special health problems of women may have led white female psychiatrists to over-emphasize the contribution of psychosocial issues to psychiatric disorders in white females.

### **Reporting of distress and depression**

Weissman and Klerman (1977) questioned whether females are under more stress, or perceive life events as being more stressful. They concluded, from their review, that females do not experience more stressful life events nor judge life events as more stressful, but noted that life event scales focus on acute stressors and do not take chronic stressors into account. Since then, other authors (Briscoe, 1982; Gove & Hughes, 1979; Verbrugge, 1985) have also concluded that there is no evidence that women report a greater number of stressful events than men. However, Weissman and Klerman (1977) did find that females report more symptoms, particularly those of anxiety and depression. Briscoe (1982) concurred with this finding and noted that some women, particularly married women, reported higher levels of depressive symptoms than men, also that wives tended to worry more about their own and their family's health (including their own) than do their husbands (Briscoe, 1982). This finding may mean then that while females regard life events in the same fashion as men, they are concerned more by stressful life events in their immediate environment than their husbands. Briscoe also found that women, regardless of marital state, tend to see their symptoms as being more severe, and reported that women are more likely than men to express positive, as well as negative affect, and that some subjects who have higher levels of negative affect also report more positive affect. The possibility that expression of a wider range of affect may be a reflection of a sex-linked personality trait is rarely considered.

Several studies have suggested that social desirability factors do not prevent men from disclosing depressive symptoms, with men being found to be equally as willing as women to confide feelings of depression both in public and in private (Clancy & Gove, 1974; King & Buchwald, 1982; Bryson & Pilon, 1984). There is some difficulty in generalising from the latter two studies as they comprised university students, presumably a relatively socially homogeneous group with values that may promote self-expression. Tousignant et al (1987) found that married women did not confide more symptoms than married men but did report a higher frequency of minor rather than major symptoms.

Angst (1984c) has called attention to sex differences related to the time period under review. He noted an equal sex ratio for point prevalence of depressive episodes, but a female excess when examining 12-month prevalence in the same group. He attributed this effect to male sex-role stereotyping, leading to males being more likely to forget episodes over time. He suggested that sex differences evident using RDC and DSM-III categories disappeared if different cut-off levels for the number of symptoms per depressive episode are applied to males and females. This device was an attempt to accommodate the fact that women tend to complain of more symptoms than men for an equivalent level for impairment.

### **Help-seeking patterns**

There have been consistent reports of women being more likely to present themselves for help (both at general practitioner and outpatient facilities) and to receive psychotropic medication (Belle, 1980; Verbrugge, 1985). Help-seeking patterns were also influenced by other factors, such as marital state, social class and level of education (Ingham & Miller, 1983), so that widowed, divorced and separated women, and those with less education of either sex were more likely to seek help from a general practitioner than from a specialist. Blacker & Clare (1987) reviewed studies of psychiatric disorders presenting in general practice and noted that females had up to five times the prevalence (per number of consulting patients) of depressive disorder. They reported that a diagnosis of a depressive disorder could be made for 8-10% of consecutive general practitioner consultations, with a preponderance of women. They also stated that general practitioners tend to under-diagnose depressive disorder in the very young, the very old, men in general and the better-educated, and that a general practitioners' notion of what constitutes a depressed 'case' "is dependent upon previous experience with large numbers of young depressed women".

Thus, there may be a sizeable group of depressed subjects in community studies who have not presented for formal psychiatric treatment. Also, studies of depressive disorder in general practice groups run into the same problems of discriminating between real and

artifactual sex differences in help-seeking and diagnosis, in that women comprise the majority of subjects in most studies conducted in clinical settings, where caseness is naturally determined by the act of help-seeking.

### **Examination of 'real' factors**

The 'real' factors which could explain sex differences in depression include genetic, biological and psychosocial explanations.

#### **Genetic factors**

Sex differences for unipolar depression may be due to sex-linked transmission (Perris, 1966; Winokur & Tanna, 1969). However, most family studies find more father-son transmission than would be expected if transmission was linked to the X chromosome (Fieve et al, 1984; Gershon & Bunney, 1976). Another study (Merikangas et al, 1985) reviewed family data from 133 unipolar depressive patients (diagnosed with SADS/RDC system) and found that the male and female relatives were equally likely to be depressed, from which they concluded that sex differences in depression were not due to genetic factors. They concluded that, while membership of some families carried a higher risk of developing unipolar depression, the reason could be attributed to environmental rather than genetic factors.

Using the family data from the NIMH Collaborative Psychobiology of Depression Study (Rice et al, 1982), there was a female preponderance for primary unipolar depression in older age subjects which diminished in younger age subjects. The mean age of onset of depression also fell from 38 and 40 years for mothers and fathers of subjects to 20 and 19 years for daughters and sons of subjects, demonstrating earlier age of onset in younger cohorts, or diminished recall in older cohorts. The presence of depression in a mother was more predictive of depression in offspring of either sex than depression in a father. There were no sex differences for bipolar disorder or secondary depression.

Winokur's group (Winokur, 1985b) have also shown an increased risk of unipolar depression in families with a strong history of depression and alcoholism and claims to have found a genetic link between alcoholism and depression, with a proposed link between depression and the female chromosome, and between alcoholism and the male chromosome.

### **Biological factors**

The hypothesis that women are vulnerable to changing levels in hormones (both through the menstrual cycle and at menopause) has been considered extensively. Nolen-Hoeksema (1987) has reviewed research considering premenstrual hormonal fluctuations on mood and found many methodological flaws, including use of retrospective questionnaires (which over-estimate and bias rates of depressive symptoms), lack of differentiation of depressive symptoms from more general feelings of discomfort and lack of comparability of populations in different studies. She concluded that the methodological difficulties were great and had not as yet been overcome, but "the evidence suggesting that biochemical fluctuations lead to mood changes is indirect, open to multiple interpretations, and contradicted by an equal amount of negative evidence".

She also reviewed studies of post-partum and peri-menopausal depression. There is evidence that the first six months post-partum is a time of increased risk for psychotic illness (Kendell et al, 1987) and periods of mild dysphoria, the so-called 'post-partum blues' (Stein, 1982), but there is little evidence that the birth itself leads to an increase in rates of non-psychotic depression. Research into rates of depression in the post-partum period is complicated by the co-existence of such symptoms as tiredness, irritability and weepiness which will generate high levels of false positives if self-report scales are used with cut-off levels designed for the general population (O'Hara & Zekoski, 1988). While there is a discrete group of women for whom post-partum depression persists for three to six months (Cox et al, 1982; Cox et al, 1984). Two other groups, O'Hara et al (1984) and Atkinson and Rickel (1984) found that many of the women who were depressed post-partum had already been depressed prior to the birth. O'Hara has called for more use of non-pregnant, age-matched

controls and for longitudinal designs in post-partum depression studies. His own study using this methodology has found no differences in rates of depression between women followed over a pregnancy and their matched controls.

The high rates of depression in women aged 35 to 45 years has led to interest in the effects of menopause. Recent research, however suggests that depression during menopause is not a distinct entity and that the incidence is not increased in the peri-menopausal years (Weissman & Klerman, 1977). Hallstrom and Samuelsson (1985) found no link between depression and naturally occurring menopause, although there may be a link between depression and surgically-induced menopause (McKinlay et al, 1987). The latter study of 2,500 post-menopausal women concluded that social factors played a much greater causal role in depressive disorders than endocrine factors.

In a review of sex differences related to biological factors putatively linked to depression, Halbreich et al, (1984) reported that depressed females have higher levels of 5 hydroxy-indole-acetic acid (or 5HIAA, the main metabolite of serotonin) and homovanillic acid (or HVA, the main metabolite of dopamine) in their cerebro-spinal fluid (CSF) when compared to depressed men. They also reported that higher plasma cortisol levels found in females are linked to age, with no sex differences after females reached menopause. Another group (Hunt et al, 1989) found that neither age nor gender had an effect on plasma cortisol levels but that plasma dexamethasone levels increased with age. A study of sex differences in cortisol, ACTH and prolactin responses to 5-hydroxy-tryptophan (5-HTP) in healthy controls and depressed patients (using ICD-9 diagnoses) found that healthy men have significantly higher levels of serum cortisol than healthy women, but that in melancholia the situation was reversed, so that women have significantly higher levels than men, with increased cortisol, ACTH and prolactin responses to 5-HTP. They concluded that "the central serotonergic regulation of ACTH and prolactin is significantly different between the sexes and between healthy controls, minor (i.e. neurotic) depressives and severely depressed patients". They found no differences for pre- and post-menopausal females and therefore considered all



females as one group. They stated that their findings only applied to melancholia, as responses for healthy controls and patients with an ICD-9 diagnosis of neurotic depression had similar hormonal responses that were quantitatively different to those of melancholics.

The NIMH Collaborative Program on the Psychobiology of Depression noted the small sample size used in many of the studies investigating biogenic amine metabolism. The authors (Koslow and Gaist, 1988) used a group of 132 subjects, which they considered large enough to account for sex and age effects. They, too, found a complex relationship between age and sex. Pre-menopausal depressed women had lower CSF levels of HVA than both age-matched depressed men and controls, while in the over-50 age group, depressed subjects of both sexes had lower levels. In the case of 5-HIAA, levels were higher in females across all age groups and 3-methoxy-4-hydroxyphenylglycol or MHPG levels increased with age in both sexes. These findings highlight the need to control for age and sex effects in this type of research.

Nolen-Hoeksema (1987) stated that "it would be premature to conclude that hormonal fluctuations have no effects on mood in women, because many studies of hormones and moods have serious methodological flaws....and the biological explanations of sex differences do not explain the absence of sex differences in certain subgroups. Psychosocial factors, such as the supportiveness of the Amish culture or the greater impact of a spouse's death on men than women, more convincingly explain the variations across groups in sex differences in depression".

It may be that biological factors are more directly implicated in endogenous depressive subtypes in the older age group (i.e. both older men and post-menopausal women), while psychosocial factors are more relevant in the younger age group and neurotic/reactive or non-melancholic depressive subtypes.

Gitlin and Pasnau (1989) listed five separate issues when examining psychiatric syndromes linked to female reproductive physiology "which interact to obscure research findings". (i) Consideration of psychodynamic aspects of sexuality and reproduction allow the potential for feminine sex-role and identity to be entwined with expressions of psychiatric disorders. (ii) As no clear association has been demonstrated between biological-hormonal abnormalities and pre-menstrual syndrome or post-partum depression, the focus should change "to examining biological vulnerability factors or potentially abnormal responses to normal biological changes". (iii) The introduction of 'late luteal phase dysphoric disorder' in DSM-III-R, which has introduced some reliability to the concept of timing of pre-menstrual syndromes, but this does not establish validity for the concept. (iv) The introduction of this disorder means that an event which is seen as a normal occurrence in many cultures is being medicalised in Western, industrialised culture. (v) The need to identify pre-existing psychiatric disorders, so that pre-menstrual and post-partum exacerbation of ongoing psychiatric disorder can be distinguished from discrete episodes that are clearly related to the reproductive cycle only.

### **Psychosocial factors**

There is a complex interaction between sex, social class, level of education, life-style, and employment status when considering risk factors to the onset of unipolar depression and help-seeking patterns related to such disorders. These issues have more often been studied by sociologists, who are more likely to examine psychological distress or depressive symptoms than categorical depressive psychiatric disorders. This results in data not easily comparable across the two fields. Some of the reviewed studies report risk factors for depressive diagnostic categories and others report depressive experience in terms of self-report symptom measures.

Charney and Weissman (1988) have reviewed risk factors for unipolar depression, which include: being female, particularly between the age of 24 and 45; having a family history of depression or alcoholism, poor quality early parenting, recent negative life events,

particularly recent separation or divorce, lack of a confiding intimate relationship, marital discord, and having had a baby in the preceding six months.

They also reviewed risk factors associated with depressive symptoms which include: being a young woman or an older man, being of lower socio-economic class, being of non-white racial origin, recent divorce or separation.

This section will give an overview of psychosocial risk factors to the onset or maintenance of depression. Potential risk factors to outcome are not considered here.

i) Age

Jorm's (1987) review of studies detailing age of onset in depression indicated an increase in rates of depression in females aged 25 to 40 years, while the male rates remained stable.

Age cohort studies have noted an earlier age of onset and higher rates of major depression for females in cohorts born since 1936, with one study finding no change in sex ratios (Weissman et al, 1984). Others (Hagnell et al, 1982; Rice et al, 1984; Klerman et al, 1985) have noted sex ratios for rates of depression approaching unity in more recent decades for younger cohorts (i.e. age less than 25 years).

Amenson and Lewinsohn (1981) used a self-report measure (CES-D) and case-finding instrument (SADS) to diagnose depressive episodes in a general community sample of 998 subjects. They concluded that there were no sex differences in numbers of new cases, age of onset or average duration of episodes, but that women, once they had experienced an episode of unipolar depression, were more likely to have further episodes than men.

ii) Drug and alcohol use

Weissman and Klerman (1977) had raised the possibility of the sexes using different coping styles when depressed, with women going for help and taking medication, and men using alcohol or being involved with law enforcement agencies. Since then, studies using the DIS have reported a lack of sex differences when categories of depression and drug and alcohol abuse are combined, supporting the concept of depression being masked or dealt with by differing coping strategies. For instance, a study using DIS/DSM-III diagnoses in Mexican-Americans and non-Hispanic white residents of two Los Angeles communities found that non-Hispanic whites had higher six-month (Burnam et al, 1987) and lifetime rates (Karno et al, 1987) of drug abuse and major depressive episodes, particularly in young women. Alcohol abuse and dependence were common in both ethnic groups for men of all ages, while antisocial personality was predominantly a disorder of young men of both ethnic groups. A study of 260 males and 241 females with drug and alcohol problems in an addiction unit (Ross et al, 1988), interviewed using the DIS, found no difference in current prevalence of affective disorders, total number of psychiatric disorders (when DSM-III exclusion criteria were ignored) or scores using the Beck Depression Inventory (Beck et al, 1961) or General Health Questionnaire (Goldberg, 1972). They did find sex differences in some disorders, with females being more likely to attract a diagnosis of anxiety, psychosexual disorder or bulimia, and males more likely to have a diagnosis of antisocial personality or pathological gambling. There were sex differences in patterns of drug abuse, with females more likely to abuse prescribed drugs and males to abuse cannabis and tobacco. Married patients of both sexes had the lowest risk of major depression, as did the more highly educated. Unemployment increased the risk for both sexes.

iii) Urban and rural dwelling

Rates of depression have been found to be higher in urban dwellers when compared to rural dwellers (Crowell et al, 1986). Blazer et al (1985) reported data from 3798 adult community residents in Piedmont, North Carolina, as part of the ECA study and found that rates of major depression were nearly three times higher in urban than in rural counties after

controlling for age, sex, race, marital status and residential mobility. These differences were greater in the 18 to 44 age group and most pronounced for young women, so that the likelihood of having an episode of major depression was influenced by sex, being a mother with small children, experience of recent stressful life events, socio-economic class (being greatest in upper middle class), absence of a confidant (giving nearly four times the risk), and urban dwelling (Crowell et al, 1986), while rates of alcohol and drug abuse were higher in rural and less well-educated subjects (Blazer et al, 1985). They concluded that rural dwelling was more of a buffer to depression for young women than for young men and raised the possibility that young women were more sensitive to their environment than young men. They did consider the possibility that alcohol and drug abuse may be alternative expressions of depression in young men but did not discuss another possible explanation, that rural living may encourage drug and alcohol abuse in young men.

iv) Marital state

For unipolar depression, marital state influences rates of depression differently for each sex. Rates for single men are higher than for married men while the converse is true for women; with married women reporting higher rates than married men. Recently separated women have the highest rates of all (Radloff, 1975; Gove, 1983). However, when employment status is added to the risk factor equation, married men who are unemployed or employed in professional careers are found to be more depressed than married women in comparable situations (Radloff, 1985). Employment status has some differential sex effects, in that women with unemployed husbands experience high levels of depressive symptoms (Radloff, 1985), while in another study, wives' employment status did not have a reciprocal effect on levels of depression in men (Cochrane and Stopes-Roe, 1981).

There is also a relationship between sex role, marital state and depression. Pearlin (1975) related that "role disenchantment" increases in women with the number of children at home and, the younger the child, the greater the disenchantment. Hafner (1986) concluded that marriage and parenthood protected men from depression but "the balance of evidence

suggests that it is the combination of marriage and lack of employment outside the home that is most potent in causing depression in married women". He noted that the presence in the home of children younger than 12 years was strongly associated with depression in married women "because it prevents married women from working outside the home". He also concluded that both men and women who adhered to rigid sex role stereotypes were more vulnerable to marital difficulties, and the onset of depression; and that they tended to cope with such difficulties by becoming even more rigid, which in turn impacted on their marital relationships, so escalating the process.

Briscoe's study of sex differences in psychological well-being in general practice attenders established the lowest levels of psychological well-being in full-time housewives and older single men (Briscoe, 1982).

A further complicating factor is that not only does the quality of the marital relationship affect the onset of depression, but episodes of depression have a long-term detrimental effect on the quality of the relationship. Depressed women tend to ruminate more than depressed men in response to their depressed mood, and depressed husbands have a greater ability to cause depressed mood in their wives than the converse situation (Hinchcliffe et al, 1978). One study showed that marital dysfunction was still apparent up to 48 months after depressed women had lost their symptoms (Rounsaville, 1980) but no comparable group of men have been studied. Where possible, it is therefore desirable to differentiate between personality vulnerability, increased psychological load on those individuals (not necessarily women) who have assumed a nurturant role, and the effects of an unsupportive relationship.

Bebbington's (1987) study of British national statistics for first admissions for affective disorders using ICD-9 diagnostic categories from 1982-1985 reported overall incidence rates of 36 for males and 59 for females per 100,000. The peak rates for depressive neurosis were in the 20 to 44 year age group, with affective psychosis being common at an older age (54 to 74 year age group). He found that single women had lower admission rates than married

women for all categories of disorder and that the female:male sex ratio was at its greatest for married subjects. Among the widowed, the sex ratio approached unity for every category of depressive disorder, while the rates for admission were lower in widowed and divorced men. Therefore, the relative risk of admission for affective disorder was greater for widowed and divorced men than for women in a similar category. He concluded that the study supported the finding that "marriage is a less healthy state for women than for men" but that the effects of the post-marital state are also most marked in young women possibly "because they have the most to put up with in terms of economic hardship and the care of young children".

These findings were replicated in a case registry study in South Verona, Italy (Bebbington & Tansella, 1989), which reported female to male rates of 1.3:1 for endogenous depression and 1.4:1 for neurotic depression, both rates rising with age. There was less variation by age and sex in marital groups for endogenous depression than neurotic depression if divorced and separated sub-groups were excluded.

v) Sex roles

Brown (1978) found that women who lacked a confidant (i.e. a spouse or boyfriend) were four times more likely to become depressed when exposed to a "severe event or major difficulty", but he did not investigate men, having already decided that he would have a higher 'hit rate' for depression. Vaillant (1977) followed a group of undergraduate men selected for their likely success in life and found that the ability to sustain an ongoing intimate relationship was a protective factor against onset of psychological problems of all sorts, but the study did not include women. It is unfortunate that these two interesting studies involved one sex only but, in their implicit assumptions (choosing women when higher rates of depression are sought and choosing men when success in life is being studied), they both illustrate the possible in-built biases operating in the area of sex risk factor research.

The issue of disadvantage in female social roles was highlighted in Weissman and Klermans' review (1977). They considered there were two possible social explanations: the

differential in factors such as social status, job opportunities and income between most men and women; or factors involving sex- role behaviour and learned helplessness. Radloff (1985) has investigated the relationship between sex, marital state and employment issues, and suggested that there may be both biological and social factors operating in the case of susceptibility and precipitating variables, with learned helplessness or a "helpless style of coping" being seen as a predisposing factor.

Weissman and Klerman's review (1977) concentrated on potential risk factors in females and only briefly discussed the issue of the methods males used to deal with stress. The function of the male sex role was hardly considered either as a protective factor against depression or as a factor leading to increased rates of social deviance which may act as an alternative to expression of depression in males.

Gove (1984) defined 'nurturant' roles (where one's role requires one to be responsive to the needs of others) and 'fixed' roles (where one has role obligations that are not easily rescheduled). He postulated that the nurturant role (usually adopted by women) is associated with high social demands, lack of privacy and high levels of distress, which may be manifested as anxiety, tiredness and lack of energy. By contrast, the fixed role (usually adopted by men) is associated with good mental health and lower psychological morbidity. He suggested that the fixed and nurturant roles complement each other, and that much of the female excess in psychiatric morbidity could be explained by the supposition that females were more likely to occupy nurturant roles. He also hypothesised that within this role, women are more likely to construe distress as a problem, and that housewives in particular (because of lack of fixed roles) find it easier to spend more days in restricted activity (e.g. more days in bed) and to partially occupy a sick role. He stated there was a higher association between fixed roles and being male, and, according to the fixed role hypothesis, "persons with fixed role obligations are less likely to define themselves as ill". Their obligations mean that it is harder for them to adopt a partial sick role but they are more likely to adopt a total sick role. He also reviewed the effect of the presence of children on their parents, noting that the



presence of pre-school children led to the greatest amount of marital strain and depression. He reiterated a former statement that "it should be noted that marital satisfaction is a powerful predictor of mental health, and that this relationship is much stronger for women than for men".

Nolen-Hoeksema (1987) examined sex differences in coping styles in depressed subjects and reported females as more likely to seek consolation but that they use less active coping styles than males. She hypothesised that females' coping styles may lead to further episodes of depression while males recovered more quickly or were less likely to have subsequent episodes. This argument is interesting but possibly flawed as it does not explain differences in incidence, but rather seems to explain possible differences in resilience and speed of recovery from a depressive episode. She also assumed that the more active style of coping is superior for both sexes.

Turner (1987) studied marriages in which one spouse was physically disabled (most commonly with heart disease and arthritis) to examine sex differences in depression (measured by the CES-D) and to consider the effects of exposure and vulnerability to stressful experience. Separate analyses were undertaken for the 18 to 44, 45 to 64 and over-65 age groups. There were sex differences in exposure to stressful experiences in the under-65 age groups, with women experiencing more stressful events, because they were in contact with a wider social network and more aware of adverse experiences within that network. He offered a social explanation for sex differences in depression, namely, that women are more responsive to the negative experiences of others, particularly their spouses, so rendering themselves vulnerable to depression, while the converse did not apply. Also, while there were no sex differences in immediate recall, over longer time-periods women were more likely to recall negative life events than men. There were sex differences in depressive symptom scores in the 45 to 64 and over-65 year age groups, with higher rates in women. He then considered depressive symptom scores in relation to events happening to subjects and to their 'significant others', together with employment status. He found that unemployed women were

more exposed, and more vulnerable to life events occurring both to themselves and 'significant others' than employed women, who were protected by their employment status to the same extent as men. Unemployed men were substantially more vulnerable to stressful events occurring to themselves than employed men.

In 1987, Weissman reviewed the epidemiological findings relating to major depression in the ten years subsequent to the original paper (Weissman & Klerman, 1977). She stated that the major achievements were the instigation of specific diagnostic criteria with consequent improvement in reliability; standardised methods for assessing symptoms and signs by direct interview or family history; and the use of large family genetic studies to determine rates in relatives of probands. She drew from the ECA Study data to note that major depression was most common in the young and "women aged 18-44 have by far the highest rates". She noted the age-related increase in rates of depression in younger cohorts and suggested that these findings were real, rather than upholding such explanations as a memory effect, changes in labelling of illness or selective survival.

She found that the lowest rates of depression occurred "in men and women who are married and getting along with their spouse, and a 25-fold increase in rates for either sex in the presence of an unhappy marriage." She discounted any effect of education, income, race or social class but upheld the finding of lower rates in rural areas.

### Summary and conclusions

A broad range of putative factors contributing to the onset and maintenance of depression have been considered, with a particular focus on those leading to potential sex differences in depressive experience. Factors hypothesised as leading to 'artifactual' sex differences include differences in (i) patterns of help-seeking and (ii) expression of distress and depression and (iii) effects of sex biases in wording of measures and diagnostic criteria.

Factors hypothesised as leading to 'real' differences include those due to (i) genetic (e.g. due to sex-linked genetic transmission), (ii) biological (e.g. due to effects of female sex hormones on mood), and (iii) psychosocial factors, including age, patterns of drug and alcohol abuse, coping styles when depressed, marital state and sex roles.

Sex differences in depression experience (for unipolar depression as a categorical diagnosis or depressive symptoms) are contingent on variables such as age and marital state. This finding suggests that sex differences in rates of depression are unlikely to be accounted for by purely biological factors and encourages pursuit of psychosocial factors that are causal or interactive with biological factors.

## CHAPTER FIVE

### THE DEVELOPMENT OF THE INTIMATE BOND MEASURE

#### CONTENTS

Introduction  
Consideration of existing measures  
The development of the Intimate Bond Measure  
Test construction and initial sample  
Analysis of data from initial sample  
Reliability studies  
Validity studies  
Discussion  
Application of the IBM

#### Introduction

The discussion in the last chapter highlighted relationships between sex, marital state and depression. While marital state can be presented simply as a socio-demographic variable, an investigation of risk factors to depression may be enhanced by inclusion of a measure of the quality of the relationship rather than merely the structural issue of whether subjects are married or not. This chapter will describe the development of a self-report questionnaire of qualitative aspects of intimate relationships after considering several measures of 'intimacy' and 'adjustment' currently being used to describe marital or other intimate relationships.

#### Consideration of existing measures

On the broadest level, there has been considerable research interest in the relevance of social support systems and, in particular, intimate relationships in influencing an individual's resilience to adversity and the onset of neurotic decompensation. While social support has been postulated to have a non-specific positive effect, Cassel (1976) has suggested that the presence of adequate social support has an additional buffering effect on the individual. Henderson et al (1981) suggested that perceived adequacy may be more important than quantitative aspects of the social support network in neurotic subjects, and their research highlights the need to examine subjective requirements of the individual when assessing social networks.

The importance of lack of intimacy as a risk factor affecting the onset and course of neurotic depressive disorder has been suggested in a number of studies. Brown and Harris (1978) isolated four vulnerability factors to depression in a non-clinical female sample, one being the lack of a confiding intimate relationship. In a replication study in Alberta, Costello (1982) examined risk factors to depression in a female sample, and reported that a lack of intimacy with spouse, cohabitant or boyfriend increased the risk of depression. In fact, Harris and Brown (1985) noted that an intimate confiding relationship with a spouse has been described as protective against depression in nine of the ten relevant cross-sectional studies of female subjects. Coyne (1987) has persistently advocated the need to clarify the complex relationship between the quality of the marital relationship and both the onset and perpetuation of depression. In his review, he highlighted some special areas of research interest, namely the effect of an intimate as a protective factor against the onset of depression, or as a predictor of negative outcome in depression, the relationship between hostile criticism and relapse, and the vulnerability of spouses and children of depressed persons. He concluded that research in the area emphasises either the biological or interpersonal approach, rather than the relationship between the two.

The reparative capacity of social support has been documented. Quinton et al (1984) noted that, while selection of one's mate is influenced by one's own experiences, selection of a non-deviant spouse who provides emotional support and good living conditions produces a significant protective effect against further emotional difficulties, even in a setting of adverse early experience. Parker and Hadzi-Pavlovic (1984) drew a similar conclusion after studying women bereaved of mothers early in life. In that study, any diathesis to depression established by earlier parenting appeared capable of significant modification by caring characteristics of the spouses.

Various strategies have been used to measure the quality of intimate relationships. The Interview Schedule for Social Interaction (ISSI), a semi-structured interview developed by Henderson et al (1981) as a global measure of social support, has been judged as limited in its

capacity to measure confiding relationships as only one question addresses that issue (O'Connor and Brown, 1984). Brown and Harris (1978) developed a semi-structured interview approach to assess psychosocial factors, but only included one question that inquired specifically into the presence of a confidant to whom the subject might turn to discuss a problem, and the frequency of such contact. This approach was modified by Costello (1982) who separated the question concerning intimacy with spouse from that assessing intimacy in confidants. The validity and utility of these approaches remain to be further established.

Concurrent with this increasing sophistication in investigation of social networks, there have been attempts to classify and measure dimensions of marital relationships, including marital "happiness", "success", "stability" and "adjustment". The Locke-Wallace scale (1959) grouped key items from existing scales in an attempt to measure and predict marital adjustment. The scale was influential for the next two decades but has been validly criticised as being partly a measure of conventionality and social desirability (Laws, 1971). Despite its shortcomings, this measure is still being used and is also the commonest yardstick for concurrent validity (Spanier, 1979).

Spanier (1976) developed a widely-used "dyadic adjustment scale" which assesses satisfaction, cohesion and affectional issues. It focuses on "adjustment" as a general indicator of marital quality and the test is offered as an aid to uncovering problem areas in close relationships.

A further impetus to the development of measures of marital function was provided by British studies examining the relationship between poor marital quality and neuroticism (Pond et al, 1963), and the interaction of spouses with each other where the husband was a designated patient with a neurotic disorder (Kreitman, 1964). Kreitman et al (1971) rated marital interaction in terms of 'assertiveness' and 'affection' and considered these to be independent constructs. Kreitman's group later showed differences between patient and control pairs and drew attention to the excess of "husband-dominated" marriages in the patient group (Collins et

al, 1971). Ryle (1966) developed the Marital Patterns Test (MPT) to quantify spouses' experience of affection given and received, and of domination/submission. The measure was developed initially to address a specific research question concerning the relationship of marital support and marriages with a neurotic spouse. There have been problems with the selection of items in each dimension and Birtchnell (1985) has commented on the contamination of 'affection' items by 'domination' items and there have been problems with the validity of the 'domination' items. However, the MPT has been subject to ongoing validity studies (Heins and Yelland, 1981) and refinement (Scott-Heyes, 1982), and has continued to generate interest, most recently by Birtchnell (1985).

The Maudsley Marital Questionnaire (MMQ) is a 20-item measure (Crowe, 1978) derived from the Structured and Scaled Interview to Assess Maladjustment or SSIAM (Gurland et al, 1972). The MMQ was first used to assess progress in couples involved in conjoint marital therapy. Factor analysis of the MMQ generated three scales - "marital adjustment", "sexual adjustment" and "general life adjustment". Arindell has examined the MMQ critically, undertaken further reliability and validity studies, and altered some items. The original 40-item version has now become a 20-item self-report scale (Arindell et al, 1983a; 1983b; 1985). The measure has been used to investigate the effect of marital adjustment on treatment outcome for agoraphobia (Monteiro et al, 1985) and for phobic and obsessional disorders (Cobb et al, 1980). The MMQ is less ambitious in its overall scope than the Wallace-Locke Measure but quantifies satisfaction and adjustment in the three defined areas. Arindell has suggested that the MMQ would require a measure of intimacy to ensure a complete assessment of marital quality (Arindell and Schaap, 1985).

Waring (1983) has developed an intimacy questionnaire based on the theory that "dyadic relationships can be defined by three relatively independent dimensions - boundary, power and intimacy" (Berman & Lief, 1975), with 'intimacy' being defined in terms of eight sub-scales. This questionnaire has two forms: a structured interview: the Victoria Hospital Intimacy Interview (VHII), and a self-report version: the Waring Intimacy Questionnaire

(WIQ) (Waring et al, 1981; Waring and Reddon, 1983). Schaefer and Olson (1981) developed the Personal Assessment of Intimacy in Relationships scale or PAIR, an instrument that examines five components of perceived and expected intimacy exchanged between partners. The authors point out that intimacy is a process occurring over time, and that individuals have differing needs for intimacy, which they hold "is never completed or fully accomplished".

These measures have all been developed using a similar methodology. In each case, researchers have generated items to reflect preconceived constructs thought to be important, either (i) to predict partnership problems where one partner has psychiatric problems (e.g. Ryle's MPT), or (ii) to predict outcome after a therapeutic intervention (e.g. Crowe's MMQ), or (iii) to identify problem areas and marital satisfaction (e.g. Locke-Wallace scale, Spanier's DAS) in intimate relationships.

Despite all these measures, there still seemed to be a need to define the key constructs underlying intimate relationships, commencing with a heterogeneous collection of items rather than to fit items to pre-conceived constructs.

This approach has an advantage in research terms, as too often associations between predictor and outcome variables have been spurious because the variables have not been independent of each other. For example, the use of a measure of marital satisfaction (as a predictor variable) to measure depression (as an outcome variable) leads to contamination of both sets of variables by a component of dissatisfaction. Additionally, studies investigating sex differences require a scale of equal relevance to both sexes and independent of any sex-dimorphic characteristics affected by men and women in intimate relationships.

### **The development of the Intimate Bond Measure**

The initial impetus to development of this measure came from Hinde's plea for a sound basis of description and classification in defining interpersonal relationships (Hinde, 1979). He noted the importance of specifying a limited number of unitary dimensions along which



relationships may differ and suggested dimensions of love/hate, involvement/detachment and dominance/subordinance as likely contenders.

While Hinde's suggested dimensions occur in some of the measures that have been discussed, the items were fitted to pre-determined constructs. This measure will be derived using a factor analytic technique to derive dimensions from a heterogeneous item pool. Such dimensions may or may not coincide with pre-determined constructs.

### Test Construction and Initial Sample

Items were generated from a literature review and from interviews of married subjects who were asked to describe the behaviours and perceived attitudes of their spouse. The item pool was reduced to 83 by removal of synonymous items, those found to be ambiguous in pilot studies with volunteers, and those items that appeared specific to one sex only.

As the measure was intended for general use rather than for specific psychiatric application, a general population group was sought rather than a sample of psychiatric patients. Respondents were routine attenders of 11 general practitioners and one optometrist in the Sydney area and were requested by practice secretaries to complete the questionnaire anonymously and to then "post" it in a collection box placed in the waiting room. Secretaries screened subjects according to the following inclusion criteria: married subjects, between 20 and 65 years of age, having a reasonable knowledge of English, with no evidence of dementia or psychosis, and not obviously affected by drug or alcohol intoxication or by severe illness. These criteria were also used for all the subsequent studies reported in this chapter. Each subject was asked to score their partner's attitudes and behaviours in recent times on a four-point Likert-type scale (as in Figure 5.1 which is the final 24-item version of the measure). Additionally, subjects were asked to record their age, sex and occupation, the last being to assess socio-economic status on the seven-point Congalton (1969) scale.

Forty-four of the 288 forms returned were discarded because of incomplete responses. The remaining forms (returned by 148 females and 96 males) generated the data for analysis. The mean age of the sample members was 39.7 (SD 12.8) years.

#### Analysis of data from initial sample

Principal components analysis was used with the SPSS program (Nie et al, 1975), and an oblique rotation selected. The first analysis incorporated all 83 items, and suggested 32 items to be weak discriminators with a poor distribution of responses. A second analysis entered the remaining 51 items. The unlimited factor solution was inspected as well as imposed two-factor to six-factor pattern matrix solutions. The two-factor solution appeared the most interpretable and was therefore used to generate the two sub-scales; they were labelled 'care' and 'control' and, after examining factor scores and communalities, were limited to 12 items each. Factors 1 and 2 were weakly associated (-0.36), as were the scores on the derived total 'care' and 'control' scales (-0.45) reported in Table 5.1, suggesting that high 'care' is intrinsically associated with less 'control'. Table 5.1 reports the factor loadings for the 24 items contributing to the final scales for the total sample, as well as the factor loadings derived separately for the males and for the females. The rank order of factor loadings suggested only a few sex effects. For instance, in relation to the care scale, two high-loading items for females ('is very considerate of me' and 'understands my problems and worries') contributed less strongly in the analysis of male subjects who, by contrast, loaded more highly on items 'makes me feel needed' and 'is physically gentle and considerate'. By contrast, the rank order of variables on the control scale was similar for each sex.

Table 5.2 reports data for this non-clinical group. As the minimum score for each sub-scale is 0, and the maximum 36, the mean 'care' scale scores suggest skewing while scores for the 'control' scale are more normally distributed.

Table 5.1      Factor loadings for items contributing to the  
final items in the measure

Item	Care scale		
	All	Male	Female
Is very loving to me	.81	.80	.82
Is affectionate to me	.79	.82	.79
Is a good companion	.78	.72	.80
Is very considerate of me	.75	.67	.78
Is fun to be with	.74	.71	.76
Show his/her appreciation of me	.73	.72	.72
Understands my problems and worries	.72	.66	.72
Confides closely in me	.72	.71	.71
Makes me feel needed	.71	.83	.64
Is gentle and kind to me	.70	.71	.69
Speaks to me in a warm and friendly voice	.70	.70	.69
Is physically gentle and considerate	.70	.79	.65

Item	Control scale		
	All	Male	Female
Insists I do exactly as I'm told	.71	.62	.77
Tends to try and change me	.69	.66	.71
Seeks to dominate me	.68	.77	.71
Tends to control everything I do	.67	.66	.74
Wants me to change in small ways	.63	.63	.62
Wants me to take his/her side in an argument	.61	.52	.65
Wants to know exactly what I'm doing and where I am	.60	.45	.67
Wants to change me in big ways	.59	.45	.68
Tends to criticise me over small issues	.59	.54	.62
Is clearly hurt if I don't accept his/her views	.57	.55	.59
Is critical of me in private	.55	.61	.51
Tends to order me about	.75	.84	.71

Table 5.2 Normative data in initial sample and validity studies groups

Sample	Care				Control		
	n	Mean	Median	SD	Mean	Median	SD
Initial sample							
Male	96	28.4	31.6	8.0	11.2	9.5	7.3
Female	148	27.1	29.2	8.3	9.6	7.1	8.3
Couples group (Validity study I)							
Male	25	23.6	24.0	8.6	13.8	11.0	8.6
Female	25	25.9	27.0	7.2	12.6	11.0	7.2
Non-clinical sample (Validity study II)							
	33	28.1	30.0	8.8	11.9	9.0	8.4

Table 5.3 Reliability and mood state results

Study and scale	Baseline test (Mean and SD)	Repeat test (Mean and SD)	Correlation: time 1 with time 2	t-test time 1 vs. time 2
<u>Test-retest study</u> (non-clinical group)				
Care scale	28.4 (7.2)	28.0 (8.9)	0.89**	0.65
Control scale	8.2 (7.0)	7.8 (7.1)	0.80**	0.56
<u>Mood state study</u> (depressive sample)				
IBM scores				
Care scale	24.2 (9.5)	25.6 (9.1)	0.92**	-1.50
Control scale	11.9 (8.7)	9.5 (7.8)	0.84**	2.60*
Depression levels				
Beck (n=13)	22.2 (9.1)	8.1 (5.8)	0.67**	7.6**
Zung (n=17)	55.6 (10.2)	34.6 (7.0)	0.80**	13.9**
* p < 0.05    ** p < .01				

The influences of subjects' age, sex and social class on scale scores were examined by univariate analyses and subsequently by regression analyses, but no significant associations were demonstrated. The three socio-demographic variables accounted for only 1.9% of the variance in care scores and 3.0% of the variance in control scores. Thus broad socio-demographic influences appeared to have little effect on scale scores.

The properties of the derived measure were then assessed in a series of studies.

### **Reliability studies**

Two aspects of reliability were assessed, internal consistency and test-retest reliability. Internal consistency was extremely high, with Cronbach's alpha being 0.94 for the care scale and 0.89 for the control scale (Cronbach, 1951). Such results suggest the derivation of homogeneous dimensions. To assess test-retest reliability, the measure was completed by 28 normal volunteers (mean age 34, SD 6.3 years), with the interval between presentations varying from three to six weeks. Table 5.3 shows that the mean scores did not differ on the two occasions while the reliability coefficients were very high (0.89 and 0.80,  $p < .001$  for both) supporting the reliability of the two derived scales.

As mood state is held to have the potential to influence self-report scores, the measure was completed by a group of depressives, first while depressed and after significant improvement. Severity of depression was monitored either by the Beck (Beck et al, 1961) or Zung (1965) measures, as subjects for this study were obtained from two other studies variably using those two depression rating scales. While 35 subjects were entered into this study, only 30 were included in the final analyses as criteria of minimal improvement of ten units on either the Beck or Zung scale were imposed to ensure a distinct mood change and a clinical assessment of change from 'caseness' to 'non-caseness' in each subject. All subjects had a clinical depressive disorder, most were in-patients at baseline assessment, and their diagnoses were, broadly, neurotic depression ( $n=14$ ) and endogenous depression ( $n=16$ ).

Data in Table 5.3 demonstrate a considerable and consistent decline in depression severity for the sample between occasions of testing, representing an improvement of 64% for the Beck scale and of 38% for the Zung scale. There were significant improvements in depression and high reliability coefficients (0.92 for care, and 0.84 for control). The care scale scores were not significantly altered, suggesting that fluctuations in depressed mood have little influence on the measurement of 'care'. There was a significant decrease in 'control' scores from depressed to non-depressed state, suggesting that mood state had some effect on perception of 'control'.

### **Validity studies**

As the measure was designed to assess perceived characteristics, it appeared important to assess its validity principally as a self-report measure. Thus, a heterogeneous sample of 33 non-clinical volunteers and psychiatric patients was obtained, with subjects being interviewed by two raters who asked a series of pre-determined questions aimed at eliciting the degree of 'care' and 'control' described by the subjects, but without using those actual words. These responses were rated on six-point ordinal scales. The author (rater 'A') was a consistent rater of all 33 subjects, while the second rater ('B') was one of four psychologists who volunteered their time. The inter-rater reliability coefficients ('A' and 'B' scores intercorrelated) were assessed at 0.66 ( $p < .001$ ) for the 'care' dimension and 0.70 ( $p < .001$ ) for the 'control' dimension, suggesting moderate rater consensus in assessing these dimensions. After the interview, subjects completed the measure and scale scores, and interview scores were then intercorrelated to assess the concurrent validity of the measure. 'Care' scale scores correlated 0.68 ( $p < .001$ ) with the level of 'care' judged by rater 'A' and 0.43 ( $p < .001$ ) with the 'care' judged by rater 'B'. 'Control' scale scores correlated 0.74 ( $p < .001$ ) with rater 'A' and 0.55 ( $p < .001$ ) with rater 'B' judgments of control at the interview. The higher coefficients returned against the author (rater 'A') are likely to reflect her constancy in assessing the content of the interview, so acting to reduce criterion variance.

While the aim was to develop a phenomenological measure of perceived characteristics, assessment of the degree to which the measure might reflect actual characteristics appeared important. To this end, 25 couples engaged in marital therapy were asked to complete the measure in relation to each other. One of two therapists (as an independent observer) then rated the interpersonal characteristics of the husbands and wives toward each other during therapy (using a six-point scale).

If the 'care' and 'control' scales are a valid measure of actual characteristics, scale scores should correlate with therapist ratings - assuming, of course, that the therapists were accurate raters and that the subjects behaved toward each other during therapy as they interacted generally. Raters' judgements of husbands' 'care' correlated 0.48 ( $p < .01$ ) with the wives' scores on the measure, while the equivalent examination for wives was 0.42 ( $p < .05$ ). Raters' judgements of husbands' 'control' correlated 0.51 ( $p < .005$ ) with the wives' scores on the measure, while the equivalent examination for wives returned a coefficient of 0.35 ( $p < .05$ ).

### Discussion

A relevant set of attitudinal and behavioural items was reduced using principal components analysis, to suggest two source dimensions underlying intimate personal relationships and labelled 'care' and 'control'. The mean scores for each sub-scale were internally consistent and stable over time, strongly supporting the reliability of each scale, and indirectly supporting the validity of the measure.

A depressed mood is frequently noted (e.g. Paykel et al, 1969) to influence the perception or recall of experiences, and phenomenological or self-report measures are particularly susceptible to such effects. As the measure was designed for use in studies assessing the quality of intimate bonds as a risk factor for depression, it is important to estimate the degree to which scale scores might be modified by changes in mood state. Despite significant improvement in depression levels, care scale scores were not influenced by varying levels of depression in this study. Control scores did decrease as depression levels



dropped, reflecting either an actual change in interpersonal relationships or a mood state bias. As the latter possibility cannot be discounted, research studies need to concede or control for mood state effect for IBM control scores. Waring's (1985) view that "perception of deficiencies in intimacy is not simply a result of the depressed state" requires such a qualification.

The concurrent validity of the scales was examined by the use of structured interviews with predetermined questions assessing perceived aspects of the partner's level of 'care' and 'control'. This established that the raters were in moderate agreement about both dimensions under investigation. Intercorrelation of the interview scores by the one constant rater with subjects' IBM scores showed moderately high agreement, supporting the concurrent validity of the derived measure. High correlations, however, might merely indicate that the subjects reported similarly in two contexts and similar judgements from the two raters might mean no more than a persistent response bias.

Although the priority was to develop a self-report measure of perceived characteristics of intimate relationships, it appeared important to make some estimate of the degree to which scale scores might reflect any 'objective reality'. Thus, there was an attempt to assess the validity of the measure in terms of its capacity to provide information about the 'actual' characteristics of the intimate relationship, assuming that there is likely to be some dissonance between 'actual' and 'perceived' characteristics. There are, of course, considerable problems in attempting to validate any phenomenological construct and the author is not aware of any technique that has resolved the intrinsic difficulties. The reader is referred to Spanier's (1979), Waring's (1985) and Birtchnell's (1985) reviews of the subject. Ryle (1966), in his validity study of the Marital Patterns Test, compared observations of a psychiatric social worker seeing the couple in question, his own observations of the couple in general practice and the couples' verbal self-rating, with their scores on the Marital Patterns Test. Most validity studies have, however, concentrated on comparisons with other scales, and not proceeded beyond concurrent validity.

A more commonly used strategy in extended validity studies is to compare scale scores against ratings provided by witnesses who are usually family members or associates in a specific situation. There are distinct and rarely-considered limitations to this approach. In an earlier study, Parker (1983), using a cross-over sibling study design for the Parental Bonding Instrument, noted that respondents tended to score parents in a similar way, irrespective of whether they were asked to complete the questionnaire for themselves or from the imagined standpoint of their siblings. These findings suggested that there is no such reality as the 'objective' rater and that family members may be particularly likely to introduce their own subjective distortions, weakening any validity assessment using such so-called 'independent' raters.

Crandall (1976) has reviewed studies attempting to validate self-report measures using ratings by others and has noted rather low coefficients. Here a modification of that strategy was attempted, having couples engaged in marital therapy rate each other and be simultaneously rated by their therapist, blind to their scores. Clearly, there are limitations to this technique, as a number of biases may be contributed by the partners themselves (e.g. social desirability, defensiveness, need to hurt the other, and the degree to which key characteristics will actually be expressed in sessions), while the degree to which therapists may judge such characteristics objectively remains unclear. Nevertheless, the comparison of the therapist's judgement of each marital partner correlated moderately (mean = 0.44) with the marital partners' ratings of each other, and offers support for the instrument as a measure of 'actual' characteristics of intimate relationships.

The 'care' dimension, as defined by the twelve items, reflects care expressed emotionally as well as physically, with constructs of warmth, consideration, affection and companionship. These items are nearer to Birtchnell's concept of affection and less contaminated by those aspects of control that he noted in Ryle's "affection" dimension (Birtchnell, 1985). The 'control' dimension suggests domination, intrusiveness, criticism, authoritarian attitudes and behaviours. The internal consistency analyses suggested that the

two scales are highly homogeneous, an important property of the final measure. As the factor scores and the scale scores on the two scales were negatively associated, it can be concluded that, in general terms, increasing 'control' in an intimate relationship is associated with less 'care'. Neither age, sex, nor social class had any influence on scale scores, another important property as it reduces the necessity to control rigorously for such variables in future case-control studies. The absence of any broad sex effect is somewhat surprising but the absence of sex effects has also been noted by other writers (Waring & Reddon, 1983; Ryle, 1966) and it should be recalled that items which appeared at face value to be more specific or idiosyncratic to either sex had earlier been deleted.

While dimensions of 'care' and 'control' may be central to intimate relationships, intimacy itself is defined as "a mutual needs' satisfaction" (Clinebell & Clinebell, 1970) and is construed as a bi-directional concept, with each partner measuring the intimate bond in terms of their own needs as well as the ingredients provided by the other. The IBM is principally uni-directional, measuring the degree to which the intimate is perceived as demonstrating certain attitudes or behaviours, although the form in which the items were phrased on the self-report scale must introduce a bi-directional component. While intimacy may theoretically be best effected by the combination of 'high-care' and 'low-control', it is possible that some individuals may judge a lesser degree of care as satisfactory to their needs, while others may seek or require a moderate or high degree of 'control' from their intimate. Thus, while these two factors may define key structural parameters in intimate interpersonal relationships, the judgement of 'intimacy' may still require further assessment of the recipient's needs and satisfactions (e.g. a scale which assess satisfaction with structural or functional issues within the partnership).

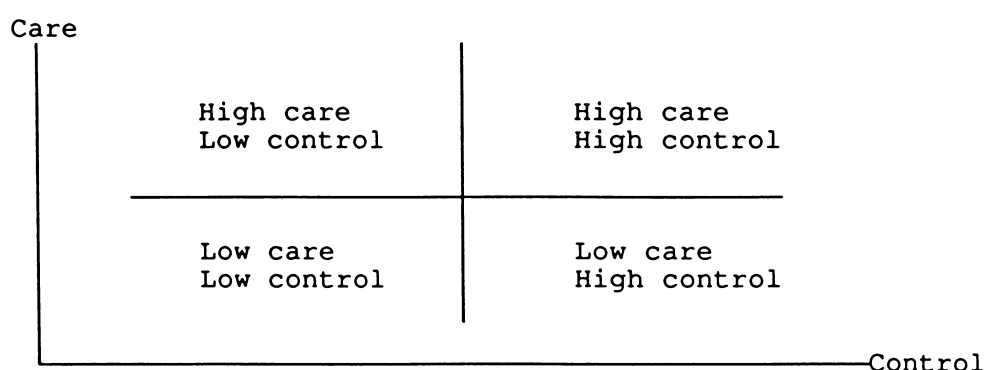
### Application of the IBM

The Intimate Bond Measure (IBM) is a simple and efficient measure of intimate relationships that may be used for risk and outcome studies. The measure (in full in Appendix 1) generates separate scores for 'care' and 'control', as has been reported in this chapter. The

scores may be dichotomised to create 'high care' and 'low care', 'high control' and 'low control' categorical variables. Essentially, one has to make a decision, after inspection of the data, as to where the cut-off between high and low levels should be. The cut-off may be at (i) the mean, (ii) the mean plus or minus one standard deviation, or (iii) median scores. The care and control subscales are dichotomised using the second method when used as predictor variables in Chapter 10 and will be described there.

With the definition of such categorical variables, it is also possible to define four broad styles of intimate relationships, as for the Parental Bonding Instrument or PBI (Parker et al, 1979; Parker, 1983) which is a measure of perceptions of early parental style with two subscales measuring 'care' and 'overprotection' for each parent. The quadrants can be labelled 'high care-low control' reflecting 'optimal intimacy', 'high care-high control' quadrant for 'affectionate constraint'; 'low care-high control' for 'affectionless control' and 'low care-low control' for an absence of intimacy (Figure 5.1).

Figure 5.1      Quadrants generated by dichotomising scores for IBM 'care' and 'control' subscales



The IBM and the PBI both measure perceptions of interpersonal relationships. The two subscales generated from the measures will be used, together with 'presence of an intimate confiding relationship' as predictor variables to depressive caseness in the cohort sample (as described in Chapter 10).

## CHAPTER SIX

### DESCRIPTION OF COHORT STUDY

#### CONTENTS

Introduction  
 Hypotheses  
 Sample selection  
 Entry assessment  
 Five-year review  
 Development of life event scale  
 Ten-year review  
 Evaluation of DIS data derived in mailed version  
 Tests of reliability of information  
 Special characteristics of the group  
 Analysis of data

#### Introduction

A group of trainee teachers was selected for longitudinal study, firstly to determine the prevalence and incidence of depressive disorder for males and females and, secondly, to examine depressive experience against a number of psychosocial risk variables. The group was selected to have certain characteristics: (i) likely homogeneity in terms of age, marital status, educational standard and career characteristics; (ii) being a middle-class group, some confounding factors, such as extreme sociopathy, poverty and high levels of unemployment were excluded, and (iii) (being at the end of a post-graduate university course), the average age for the group placed them nearer to the 'hazard age range' for onset of depressive experience.

Having effectively controlled for a number of psychosocial risk factors at base-line, it was assumed that depressive experience was likely to be similar for the two sexes at entry. The aim of the cohort study was to re-examine for sex differences over time when the two sexes would be expected to have encountered a number of stressors. Some, (such as starting work), would be common to both sexes, some others, (such as motherhood), would have a sex-specific differential. It was anticipated that career opportunities would be similar, but that men would be

more likely to avail themselves of career advancement opportunities while women would be more likely to cease work during child-rearing years. Different experiences and stressors for each sex would therefore be encountered during the mid-20 to early-30 age range when sex differences were likely to become more prominent (Jorm, 1987). If sex differences emerged, the process of serial assessment of potential risk factors would allow the relevance of aetiological factors to be specified with some precision.

In undertaking this study, a number of hypotheses were posited:

### Hypotheses

1. There would initially be no sex differences in rates of depressive experience in a homogeneous group of post-graduate university students.
2. Sex differences in rates of depression would emerge over time, with females reporting higher rates.
3. Risk factors considered to be associated with female vulnerability to depression (e.g. high interpersonal dependency, feminine sex role stereotype, exposure to home duties and motherhood) would be confirmed.
4. Risk factors considered to be associated with vulnerability to depression in both sexes (e.g. high neuroticism, low perceived parental care during childhood, low self-esteem, high exposure to negative life events, lack of perceived social support and dysfunctional intimate relationships) would be confirmed.

Table 6.1 provides a summary of the stages involved over the ten years of the study and now outlined.

### Sample selection

In September 1978, 380 students who had completed a basic Arts or Science university course, and who were then undertaking a one-year teachers' training program, were approached in class and invited to participate in a five-year research project. While it was explained that information on their depressive experience would be sought over time, they were not informed of the focus on sex differences in

Table 6.1      Flow chart of the study

---

1978	<p>Explanation of study, volunteers invited. Self-esteem rating (n=380).</p> <p>170 volunteers correctly completed base-line. data and became study group.</p> <p>29 rejected (incorrectly completed forms). 180 refused further involvement.</p> <p>Comparison of volunteers and refusers.</p>
1979	Study subjects contacted by mail.
1980	Contact made by mail.
1983	<p>Five year follow-up of 165 subjects 150 interviewed in person, of these, 133 provide corroborative witnesses.</p> <p>15 returned information by mail, of these, 14 had subsequent personal or telephone contact.</p> <p>1 subject refused further contact.</p>
1985	Subjects contacted by mail.
1988	<p>Ten year follow-up of 161 subjects, 3 males refuse interview.</p> <p>151 subjects interviewed in person, 10 returned information by mail.</p>

---



depressive experience, nor of a study of a similar group of trainees in the preceding year (Parker, 1979).

Three hundred and fifty one of these students successfully completed a questionnaire which included a self-esteem scale (Rosenberg, 1965), assessment of paternal occupation to rate social class on a four-point rating scale (Congalton, 1969), and a question assessing whether they were willing to take part in the longitudinal project. Of the 197 students expressing provisional acceptance, 170 completed and returned a mailed base-line questionnaire and were regarded as the study cohort. Scores on the self-esteem scale were compared as a brief screen to determine if the 170 study subjects (114 female, 56 male) differed from the remaining 181 students (119 female, 62 male) who had finally elected not to take part. The mean scores for the respective groups were 1.53 and 1.43, the difference not being significant ( $t=0.74$ , ns). Male "refusers" did not differ from male sample members on the self-esteem measure ( $t=1.39$ , ns) or in age ( $t=1.32$ , ns). Similarly, female "refusers" did not differ from female sample members on the self-esteem measure ( $t=0.04$ , ns) or in age ( $t=0.46$ ). Males and females did not differ in their likelihood of either refusing or joining the sample ( $\chi^2=0.06$ , ns).

### Entry assessment

At entry in 1978, male and female members of the cohort were compared on a number of measures assessing depressive experience. These included (i) state (Wilson, 1979) and (ii) trait (Costello & Comrey, 1967) depression scales. The latter scale was designed to measure a "person's tendency to experience a depressive mood". Three further measures were included to quantify aspects of personality: (iii) a dependency scale derived from the Depressive Experiences Questionnaire or DEQ (Blatt et al, 1975) and (iv) the Eysenck Personality Inventory neuroticism scale (Eysenck & Eysenck, 1964) and (v) scores from the self-esteem rating described earlier. A scale measuring two components of perceived parental environment, 'care'

and 'over-protection' in both parents, (v) the Parental Bonding Instrument or PBI (Parker et al, 1979) was also included.

Sample members completed a questionnaire giving details of their reasons for selecting teaching as a career; how difficult they thought that career was likely to be; the degree of satisfaction derived from current relationships; the quality of such relationships in five years time; any history of chronic illness; medication taken; pre-menstrual problems; duration and frequency of episodes of depression up to 1978; and details of help-seeking behaviour adopted when depressed.

#### Five-year review

In November 1983, five years after intake, a concerted effort was made to obtain follow-up data on all 170 subjects in the cohort, and 165 (97%) were successfully located. From December 1983 to May 1984, 150 (91%) of these were interviewed while, for the remaining 15 (9%), data were collected from mailed questionnaires. For most of those in the latter group who were no longer in Sydney, telephone contact was made to clarify details.

The five-year follow-up data were collected by myself (then with 12 years' clinical psychiatric experience) after receiving training in administration of the Diagnostic Interview Schedule or DIS (Robins, 1982) from Prof. Gavin Andrews and his team, who were using the DIS as the case-finding instrument.

The data were collected during a comprehensive semi-structured interview which established a framework for an informal psychiatric, family and developmental history to be taken, before administration of those sections of the DIS dealing with anxiety states and depression. Information was sought for the previous five years in terms of work experience, drug and alcohol intake, general level of health and help-seeking behaviour, amount of perceived social support (both

generally and in times of stress), and family history of depression.

Sections of the DIS covering depressive and anxiety disorders were used in full, and the opening question from the section on mania, followed by the rest of the section, if relevant. While not designed to generate diagnostic criteria for RDC minor depression, the DIS was modified slightly to allow this category to be identified and the modifications will be discussed later in the chapter.

Questions dealing specifically with depressive experience included those in the base-line questionnaire (for duration and timing of episodes); coping styles "when depressed" and a check-list combining all possible symptoms for RDC minor depression and DSM-III dysthymia.

All 165 subjects located agreed to be interviewed, although, after interview (including the DIS), one subject decided not to complete the self-report questionnaires and not to have any further involvement with the study, stating that material discussed had reminded her of previous depressive episodes that she wished to forget.

The subjects were also given the same six self-report measures used at base-line (assessing self-esteem, state and trait depression, dependency, neuroticism, and perceived quality of early parental experience). Several additional measures were completed, including the Bem Sex Role Inventory or BSRI (Bem, 1974), designed to rate characteristics deemed to reflect masculinity and femininity as two independent dimensions, along with a neutral group of characteristics reflecting social desirability; the Dysfunctional Attitudes Scale or DAS (Weissman, 1979); a measure of the importance and satisfaction derived from work and employment (Renwick & Lawler, 1978); and the Intimate Bond Measure or IBM (Wilhelm & Parker, 1988) described in Chapter 5.

The subjects were also given a check-list of positive and negative life events covering the previous twelve months and the process of developing this scale will now be described.

#### Development of life event scale

The check-list of negative events was derived from a life event scale (Tennant and Andrews, 1976) which was modified and added to in accordance with the life events deemed likely in this particular group. Positive life events were added to determine (i) whether possible sex differences in reporting were global (i.e. applied to both positive and negative events) or specific (i.e. possibly reflecting a plaintive set bias more marked for one sex); (ii) whether the sexes were having similar life experiences, and (iii) whether the presence of positive events acted to protect subjects from the effect of negative life events. The life events were compiled also with the objective of deriving likely and salient events for this group.

The measure is shown in full in Appendix II. For the positive events, subjects were asked to state whether or not events had occurred and where they answered 'yes', to rate how pleasing the events had been on a scale from 0 ("not pleasing at all") to 8 ("as pleasing as could possibly be imagined"). The range of events was designed to ensure salience for the study cohort and included such work-related items as "You were told by a number of people that you were performing extremely well at work"; such general items as "You had a holiday lasting at least a week in your own country"; and such interpersonal items as "The relationship with your partner improved dramatically". Twenty-four of the twenty-five events were the same for either sex while the item concerning pregnancy was worded differently for each sex (see Appendix II).

For unpleasant life events, a similar self-rating technique was used, with

scaling from 0 ("not distressing at all") to 8 ("as distressing as could possibly be imagined"). Items again reflected the particular issues for the cohort, with such work-related items as "You were prevented from entering, or excluded from an educational course you wished to pursue" and "You were given significantly increased work responsibilities"; such general items as "You were involved in a legal action that could have damaged your reputation to a moderate or severe degree"; and such interpersonal items as "You were told by a partner that you were no longer loved" and "Increasingly severe arguments with your parents developed". Items concerning unwanted pregnancy and abortion commenced "You or your partner...." and were appropriate for either sex (see Appendix II).

There are a number of problems which arise in relation to the measurement of life events. Firstly, events do not have the same impact or subjective importance to every person and secondly, there are problems of retrospective evaluation of events that may be influenced by current mood state and effects emerging from how others have interpreted the events. To control for variation in actual impact (for any events later occurring) and to examine for any general tendency to "catastrophize" or deny the impact of life stressors subjects were requested at baseline to predict their responses (in terms of "pleasure" or "distress" as appropriate) to the entire list of positive and negative events to generate individual predictive scores for each item. It was intended that later, when an individual subject had experienced a specific life event, then the score for that life event items would be compared with that subject's pre-event rating for the specific items and with group. In such a way, it might be possible to determine whether subjects, compared to the whole cohort, tended either to under-estimate or over-estimate the potential impact of life events.

Thus all 170 subjects were approached by mail in 1979 and 101 of the group completed these predictive ratings. For positive life events, the range of mean scores for the group was from 5.6 (out of a possible 8) for "You fall in love" to 3.2 for "A

for the group was from 5.6 (out of a possible 8) for "You fall in love" to 3.2 for "A child other than your first is born". For negative events, the range was from 5.8 for "Your partner dies", to 2.5 for "You are given significantly increased work responsibilities". In 1980, all 170 subjects were again contacted by mail and asked to state which life events had occurred in the previous twelve months and then give pleasure/distress ratings for events that they had actually experienced. Additionally, they were asked to give the same list of events to a witness who would be able to report on their previous twelve months. Seventy subjects returned the life event scales and 62 provided forms from witnesses.

Correlations between subject's and witness scores for occurrence ranged from 0.4 to 1.0 but were generally in the 0.5 to 0.8 range. For negative life events, there were correlation coefficients of 1.0 for the following items (item number in bracket): (2) involvement in a legal action; (3) involvement in a serious accident; (16) discovery that partner was unfaithful; (20) increasing arguments with parents. The lowest correlation (0.04) was for item (4) being told of poor performance at work. Remaining life events were generally in the 0.50 to 0.75 range. For positive life events, there were no correlation coefficients of 1.0, but the highest correlation coefficients were for the following items (number in bracket): (4) an overseas holiday ( $r=0.91$ ); (2) moving into a new house ( $r=0.77$ ); (1) taking up a new hobby ( $r=0.75$ ) and (17) starting a new relationship ( $r=0.74$ ). The lowest correlation coefficient was for item (14) achieving a sense of fulfilment ( $r=0.06$ ), while most of the other items were in the range from 0.40 to 0.70. These results indicate moderate agreement between subjects and their nominated witnesses in terms of the occurrence of categorical life events on the subject, with higher consistency for some of the negative life events and for the more objective events.

It had been intended to continue an annual check of life events up to the five year review in 1983, but it became clear from the poor rates of return of mailed

questionnaires that the cohort were very mobile and as the main objective was a concerted follow-up at 1983, further interim mail questionnaires were curtailed.

### Ten-year Review

In 1988, a research assistant (Ms. Curtain - a trained nurse with an Honours degree in Psychology), was appointed to conduct a ten-year follow-up of the same group. The semi-structured interview and self-report questionnaires were identical to those used in the five year follow-up. However, a simplified life event schedule (covering the previous twelve months) was used.

Prior to commencing the ten year follow-up, Ms. Curtain was trained independently in use of the DIS by Prof. Gavin Andrews' team. She was also a DIS interviewer in the Sydney centre for the WHO sponsored multi-centre comparison trial evaluating the DIS component of the Composite International Diagnostic Interview or CIDI (Robins et al, 1988) which was taking place at that time but has yet to be published.

Prior to each interview in the cohort study, Ms. Curtain was given some background information concerning the subject: namely, marital status, number of children, circumstances under which the previous interview was carried out, and occasionally, some nuance such as an unusual hobby, so that there was some continuity and point of communication on meeting. However, no information was given to Ms. Curtain concerning the author's previous estimate of depressive experience and personality, to allow independently derived judgements on such matters to be compared.

In 1988, contact was made with all 164 subjects who had agreed in 1983 to continue in the study. Three males then requested not to be interviewed, although two of these returned the self-report questionnaires. Of the three, one subject had

just been admitted to a psychiatric clinic with a diagnosis of schizophrenia, the other two males were physically ill (one was recently diagnosed with leukaemia, the other had chronic renal failure and had suffered two failed renal transplants and recommenced a haemodialysis programme but was having medical complications). In all three cases, it was felt that the subjects' requests for privacy should be met. All three would have been expected to have rated as significantly depressed but for at least two, episodes may well have been excluded by the DIS algorithm as being due to medical causes. Two of these males had previously been allocated to the category of minor depression, while one had been free of depression. None of the three had attracted a diagnosis of major depression, dysthymia or anxiety disorder.

One hundred and fifty one subjects (101 females and 50 males) were interviewed personally and 10 subjects (7 female, 3 male) were issued with a mailed version, because of geographical remoteness. This version (like that used in 1983) contained the same questions as used in the semi-structured interview in the identical format and a specially prepared mail version of the DIS.

#### Evaluation of DIS data derived in mailed version

The DIS was originally designed for direct interview. Since then, a version has been developed for computer use which was discussed in Chapter 2 (Blouin et al, 1988). The mailed version used in 1983 and 1988 followed the same format exactly as the interview version, with questions being asked as to whether symptoms had ever occurred, and if so, when. On return of the forms, the subject was telephoned to ensure complete understanding of the task, and in the event of positive responses, the subject was asked for the number of symptoms that had occurred episode by episode to assess concordance of reporting.

To evaluate the acceptability and accuracy of such non-interview derived data, in 1983, 10 subjects (3 female, 7 male) were requested to complete the mailed



version of the DIS assessing anxiety and depression symptoms, and were subsequently interviewed in person. For 9 of the 10 subjects the judgments made at interview were the same as those made from the mail-generated data, although there was a tendency for subjects to report more symptoms at interview. For one subject, an episode of major depression was elicited at interview that had not been otherwise reported. Overall, it seemed that collection of data by mail was likely to lead to a slight under-representation of episodes of depression in those who were possible 'cases'. Nevertheless, in order to achieve a high response rate for those entering the study 5 years earlier, and because any bias was small and affecting only 15 of the sample, the mailed version was included where relevant but, any depressive symptoms recorded in mailed questionnaires, were clarified by telephone contact.

#### Tests of reliability of information

At base-line, there was no attempt to corroborate information gathered.

At the five-year follow-up, after the personal interview each subject was asked to nominate another person who had known them well for a number of years and who might be able to provide an independent account of the subject's depressive experience. Consent and a completed corroborative interview were achieved for 133 of the 150 (88.7%) subjects interviewed in person, and therefore 80.6% of the whole cohort. Where and when possible, the nominated witness was interviewed immediately to avoid discussion and "priming" by the subjects, and this occurred for 56 of the 133 corroborative interviews. If the person nominated was not available, the subject was asked to inform that witness that there would be telephone contact in the next few days but not to describe the exact nature of the information to be sought. This procedure was adopted for the remaining 77 subjects. On a number of occasions, more than one witness was sought to clarify details.

Whether contact was in person or by phone, a similar approach was used. The objectives of the research were briefly explained (but without reference to the issue of sex differences) and the following question was put: "Has A, in the time that you have known him/her, ever experienced an episode of depression lasting at least two weeks, when he/she seemed depressed or sad or behaved very differently from normal or gave you cause for concern?". If the answer was negative, the informant was prompted once more "are you sure that....?" and, if again negative, no further exploration occurred. If the answer was positive, the informant was then asked: "Could you tell me when the episode(s) occurred and something about it (them)". Information was sought as to whether there was a persistent and qualitative difference in mood and behaviour, the timing of onset and duration of episodes, and the impact of the episode on the subject and others. If the informant volunteered details of possible causes for such episodes, this was noted but not otherwise pursued. Information concerning timing and severity of episodes was compared by the interviewer at that time to that given by the subject without the informant being aware of what the subject had reported. If there were discrepancies (e.g. a single episode with identical precipitating life events was noted by both but was dated inconsistently), further questions were asked for clarification. If on further questioning, it was clearly the same episode and both were able to give adequate explanation of the disparate timing, agreement was noted.

At the ten-year follow-up, ten interviews (using the DIS) were undertaken, to establish inter-rater reliability of the two study raters, five prior and five subsequent to the commencement of the formal study. Of the ten subjects, four were patients who had presented with symptoms of depression, and six were non-clinical, hospital staff who had no connection with the study.

Complete interviews (involving semi-structured interview and DIS) were also carried out on twelve study subjects. In both groups, Ms. Curtain and the author

Complete interviews (involving semi-structured interview and DIS) were also carried out on twelve study subjects. In both groups, Ms. Curtain and the author alternated as the interviewer, with each scoring responses independently. This is the same strategy as was used in PSE and DIS inter-rater reliability trials mentioned in Chapter 2. Results showed high inter-rater reliability (100% agreement) for both the studies across all diagnoses (see Table 6.2).

Table 6.2      Inter-rater reliability studies using the DIS to generate RDC categories, carried out in 1988

Study group	Numbers of subjects in each diagnostic category	
	Diagnostic category	
	1983 Rater KW	1988 Rater RC
<u>Volunteers (n = 10)</u>		
RDC probable major depression	3	3
RDC minor depression	3	3
RDC intermittent minor depression	0	0
No episode of depression	4	4
<u>Subjects in study cohort (n = 12)</u>		
RDC probable major depression	3	3
RDC minor depression	5	5
RDC intermittent minor depression	0	0
No episode of depression	4	4

### Use of the DIS to generate diagnostic categories

Some specific comments concerning use of the DIS are required. While DSM-III uses diagnostic criteria for major depression and dysthymia, it merely offers a brief description in defining a minor depressive disorder such as 'adjustment disorder with depressed mood', so allowing considerable subjectivity in rating. While the DIS is not designed to generate diagnoses for minor depressive disorders, the standardized questions generate material readily encapsulated by the RDC definition of minor depression. A 'case' of RDC minor depression was allowed if there had been a depressive episode lasting at least 2 weeks and possessing two of the eight DSM-III symptoms used for major depression, with RDC functional impairment criteria also fulfilled. This allowed for inclusion of subjects to the category of RDC definite minor depression, who reported a depressive disorder with 2-3 of the 8 possible symptoms, with a duration of more than two weeks but less than two years. SADS/RDC provides 8 further symptoms such as "self-pity, needing reassurance or help from somebody, brooding about unpleasant events that have happened" and a category of 'probable' minor depression for an episode of one week's duration, so that the criteria used here were stricter than those in SADS/RDC system.

### Special characteristics of the group

The subjects in the study are notable for their level of intelligence, degree of cooperation and accessibility at follow-up. This was facilitated by the group's clear commitment at the commencement which involved provision of a long-term contact address. Thus, they are not comparable with a randomly selected general population group, as the hypotheses were predicated on the selection of a socially homogeneous group.

Also, the DIS was administered at the end of a semi-structured interview, rather than in isolation as had been the case in many studies, including the ECA

study. This enabled the gathering of a wide range of background information prior to administration of the DIS, with the possible advantage of prompting higher recall of depressive symptoms and episodes than in a 'stand-alone' DIS interview. A timeline which noted the number and timing of episodes along with significant life events was also employed at the end of the interview.

The diagnostic categories generated by DIS allowed calculation of prevalence over lifetime, and new case rates over the discrete five-year periods of the study.

'Caseness' was determined using both RDC and DSM-III systems. Subjects were included as RDC cases if they had experienced episodes of major depression (definite or probable), or definite minor or intermittent minor depression, so that all subjects had experienced depressive episodes lasting at least two weeks, with symptoms from a minimum of two categories with the imposition of the RDC help-seeking/functional impairment criteria. Subjects who were DSM-III cases had either experienced an episode of significant major depression, which is identical to the RDC major depression (definite and probable) or dysthymia. Subjects who were DSM-III cases had either experienced episodes of at least two weeks' duration and a minimum of four symptoms or an episode lasting at least two years with a minimum of two symptoms, again the RDC help-seeking/functional impairment criteria imposed.

With use of the SADS/L interview, the RDC does allow for one less symptom to reach threshold for RDC major depressive disorder when examining lifetime rates. As both current and lifetime episodes were being examined simultaneously, the DSM-III major depression category was used throughout.

### Analysis of data

Data were analysed using the Statistical Package for Social Sciences (SPSS/PC+ V2.0, 1988), including a computer algorithm written by Dusan Hadzi-Pavlovic, Division of Psychiatry, Prince Henry Hospital, to derive DSM-III and RDC diagnoses from the DIS-generated data.

Two-tailed t-tests were used for analysis of continuous variables. Chi square tests were used for categorical variables with Yates' correction applied if any cell had an observed frequency of less than 5. Degrees of freedom are quoted for values greater than one.

Other statistical techniques will be noted in the text where appropriate.

## CHAPTER SEVEN

### ANALYSIS OF DATA IN TERMS OF SEX DIFFERENCES

#### CONTENTS

Introduction  
 Psychosocial data at base-line, five-year and ten-year reviews  
 Perception of social support  
 Factors involved in choice of teaching as a career  
 Experience of illness and help-seeking behaviour  
 Drug and alcohol use  
 Coping styles when depressed  
 Experience of life events  
 Self-reported experience of depression  
 Self-report questionnaires  
 Self-report depressive symptom check-list  
 Perception of self as a "nervous person"  
 Summary of sex differences  
 Summary of absence of sex differences

#### Introduction

This chapter will report results from data collected at the 1978, 1983 and 1988 assessments. All data will be examined for sex differences using univariate analyses. Results for the DIS-derived diagnostic categories will be reported in the following chapter.

Of the original 170 subjects, 165 (97.1%) were successfully located in 1983, at the five-year follow-up. The data reported for base-line assessment and five-year review will therefore be for the 165 subjects (109 females and 56 males) who completed the assessments in 1978 and 1983 (see Table 6.1 for flow chart).

In 1988, 164 subjects (108 females and 56 males) were located and there are complete data for 161 subjects (108 females and 53 males), as three males withdrew from the DIS interview. Of the three males who declined interview, two completed the self-report measures, while the third provided socio-demographic and health details only.

Unless otherwise stated, the denominators for reported percentage rates are 165 (109 females and 56 males) for data collected at 1978 and 1983 assessments and 161 (108 females and 53 males) at 1988. The denominators have been adjusted to account for occasional missing values, where the number of subjects involved is less than 5% of the total and the missing values are random. The number of subjects comprising the denominator will be stated if numbers of subjects vary by more than 5% or for a specific reason (e.g. consideration of the number of subjects with partners, rather than simply the total number of subjects).

Issues concerning the reliability of data collected at five and ten year follow-up will be addressed in Chapter 9.

#### Psychosocial data at base-line, five-year and ten-year reviews

Table 7.1 summarises the socio-demographic data (for 109 females at 1978 and 1983 and 108 females at 1988; for 56 males on all three occasions). At base-line in 1978, the sexes did not differ on any of the variables examined (e.g. age, marital status, number with children, employment and social class). The mean age of the cohort (of 165 subjects) at entry was 23.4 years, and there was no difference in mean ages of the males and females. Seventy-five percent were in the 21 to 23 year age range, and only four were 35 years or older. The age range of the four subjects who had requested to withdraw from the study was 31.0 years at base-line (range 26 to 36 years) so that they were among the older members of the cohort.

Separate categories are given for married and never married subjects, but some of the 'never married' were living in de facto relationships and are reported with the married group in the 'living with partner' category. The number of subjects living with partners (whether married or not) rose markedly from 1978 to 1983, and still further by 1988, but rates remained equivalent for both sexes.



Table 7.1 Socio-demographic characteristics of the sample,  
by sex, from the 1978, 1983 and 1988 assessments

Variable	Year	Female	Male	$\chi^2$
<u>Marital state</u>				
Married	1978	23	9	0.32
	1983	58	25	0.76
	1988	69	34	0.00
Never married	1978	80	45	0.98
	1983	32	24	3.01
	1988	22	17	2.65
Previously married (divorced, widowed, separated)	1978	6	2	0.03
	1983	19	7	0.36
	1988	17	5	0.91
<u>Partner status</u>				
Living with partner	1978	23	11	0.05
	1983	66	30	0.74
	1988	79	38	0.04
<u>Number with children</u>	1978	9	4	0.00
	1983	39	17	0.70
	1988	69	33	0.04
<u>Employment</u>				
<u>Full-time</u>	1979 <sup>#</sup>	83	47	1.34
	1983	87	50	2.35
	1988	55	49	25.02***
Part-time	1979 <sup>#</sup>	21	8	0.63
	1983	13	4	0.48
	1988	33	4	9.37**
Unemployed, ill	1979	12	3	1.43
	1983	13	6	0.00
	1988	1	3	1.63
Home duties	1979	2	0	0.72
	1983	11	0	4.54*
	1988	19	0	8.95**
Social class	1978	1	4	4.98 df=3
		2	19	
		3	29	
		4	2	
<u>Age in years</u>				
	1978	23.1	23.9	t = -1.21
	1983	29.1	30.1	t = -1.51
	1988	33.4	34.0	t = 0.89

<sup>#</sup> First year of work

\* p < .05, \*\* p < .01, \*\*\* p < .001

The number of married subjects rose at each follow-up, from 21% of the females and 16% of the males at the intake assessment in 1978, to 53% and 45% respectively in 1983, to 65% and 66% respectively in 1988, with the likelihood of marrying over each interval being similar for the two sexes. Similarly, less than 10% were parents in 1978, rising to one-third in 1983 and two-thirds in 1988, the increased parenthood rate being similar for both sexes. The mean number of children rose from 0.2 for females and 0.1 for males in 1978 ( $t=0.53$ , ns) to 0.6 for both sexes in 1983 ( $t=-0.07$ , ns), to 1.3 for both sexes in 1988 (0.05, ns).

Of those subjects employed in full-time work, 81% were in teaching jobs in 1979 (the first year after teachers' college); 78% in 1983 and 73% in 1988. For those employed in part-time work in 1979, 83% were engaged in teaching jobs, falling to 65% in 1983 and rising again to 80% in 1988. From 1983 to 1988, there was an increasing number of females working part-time (generally teaching), with the likelihood of females being engaged in home duties rising from 2% (of females) in 1979 to 10% in 1983, and to 17% in 1988. When the changing work patterns for females are compared with the stable work patterns for males (no males reported being engaged solely in home duties at any time), evolving sex differences are noted, with a statistically significant preponderance of males in full-time work (by 1988), females in part-time work (by 1988) and females engaged in home duties (1983 and 1988). While males and females were becoming parents at much the same rate, the differing work patterns reflect sex differences in the effect of parenting responsibilities.

There were no sex differences in social class data based on subjects' fathers' occupations in 1978. Social class was not considered further as this factor was controlled in the choice of a sociodemographically homogeneous cohort at base-line.

---

Table 7.2 Assessment in 1978 of current and future relationship, by sex

---

Current relationship\*

	Female	Male
None	17 (15.6%)	10 (17.9%)
Superficial	9 ( 8.3%)	13 (23.2%)
Serious	16 (14.7%)	11 (19.6%)
Committed	64 (58.7%)	22 (39.3%)
	<u>106 (97.3%)</u>	<u>56 (100%)</u>

Sex difference  $\chi^2 = 9.45$ , df=3, p < .05

Anticipated relationship in five years' time\*

	Female	Male
None	0 ( 0.0%)	2 ( 3.6%)
Superficial	1 ( 0.9%)	1 ( 1.8%)
Serious	5 ( 4.6%)	3 ( 5.4%)
Committed	100 (94.3%)	49 (89.1%)
	<u>106 (97.3%)</u>	<u>56 (100%)</u>

Sex difference  $\chi^2 = 4.26$ , df=3, ns

---

\* 3 females failed to respond to the question

---

Table 7.3 "Ideal" or "close to ideal" levels of perceived social support by sex, from 1983 and 1988 data

	Year	Female	Male	$\chi^2$
<u>In normal circumstances</u>				
From partner	1983	74 (80.4%)	36 (78.3%)	0.13
	1988	76 (79.2%)	33 (80.5%)	0.01
From family	1983	75 (68.8%)	34 (60.7%)	1.08
	1988	73 (67.6%)	35 (66.0%)	0.01
From friends	1983	74 (68.5%)	36 (65.6%)	1.60
	1988	79 (73.1%)	28 (52.8%)	6.58*
<u>In times of stress</u>				
From partner	1983	75 (81.5%)	37 (80.4%)	0.12
	1988	73 (76.0%)	34 (85.0%)	3.46
From family	1983	83 (77.6%)	42 (75.0%)	0.03
	1988	75 (70.1%)	33 (62.3%)	0.83
From friends	1983	80 (74.1%)	38 (69.1%)	0.56
	1988	90 (84.9%)	26 (57.8%)	20.74***
* p < .05,    ** p < .01,    *** p < .001				

### Perception of social support

In 1978, subjects had been asked to estimate the type of relationship (e.g. 'serious, likely to be ongoing', 'committed to one partner' or 'casual') in which they were currently involved and what they anticipated in five years' time (Table 7.2). This question considered whether the females were more likely to wish to have a committed relationship in the future. In 1978, females were more likely to be involved in a 'serious' or 'committed' relationship than the males, but the substantial majority of each sex anticipating involvement in a 'committed' intimate relationship five years' hence, with no evident sex difference.

Subjects were asked in 1983 and 1988, to estimate the level of social support (reported as 'ideal', 'close to ideal', 'moderately ideal' and 'far from ideal') that they received from their partners, family and friends (i) "in normal circumstances" and (ii) "in times of stress" (Table 7.3). These questions were aimed at determining rates of perceived adequacy of social support in times of stress and data are reported in terms of those reporting 'ideal' or 'close to ideal' levels of social support. The numbers are smaller for the category of support from their partner, as some subjects (33 in 1983 and 29 in 1988) did not have a partner.

In 1983, there were no sex differences in levels of perceived social support from partner, family or friends either "in normal circumstances" or "in times of stress". In 1988, females were more likely to report 'ideal' or 'close to ideal' support from friends, both "in normal circumstances" and "in times of stress". There was a trend which just failed to reach statistical significance ( $p=0.07$ ), for males to be more likely to report 'ideal' or 'close to ideal' levels of support from their partner "in times of stress".

### Factors involved in choice of teaching as a career

In 1978, subjects were asked to rate a series of possible motivations for choosing teaching as a career. Table 7.4 reports the distribution for those subjects rating the motivating factors as 'of some relevance' or 'very relevant' to them.

Males were more likely to report being motivated towards a teaching career by exposure to an influential role model and tended to be more likely to report such motivating factors as the presence of a teacher in the family and lack of other appealing options. The opportunity of working with children and the work conditions were the most commonly endorsed motivating factors, with no sex differences.

Table 7.5 reports data for the type of teacher training undertaken. The choice of teacher training usually reflected the subject's undergraduate degree, although some of the group (particularly the males) with degrees in areas other than science and mathematics had been encouraged to undertake primary teaching. Significantly more males had gained undergraduate degrees in mathematics, science or commerce and intended to teach in these subject areas in high schools ( $\chi^2=18.82$ ,  $p < .001$ ). Females were more likely to have specialised in the humanities and social sciences, with 64% of females (as compared to 25% of males) having gained an undergraduate degree in these areas, and more intended to teach such subjects in high schools ( $\chi^2=7.03$ ,  $p < .01$ ).

There were no sex differences in those undertaking infants/primary training ( $\chi^2=1.63$ , ns) and the remainder of the cohort were involved in special education courses (teaching of the developmentally disabled and adult migrants).

Table 7.6 reports work locations in the first year of teaching in 1979, and at the time of the five-year follow-up, in 1983 ( $n=165$ ). The patterns of work location are very similar for each sex on both occasions. There were equal numbers of males and females who had never taught but were engaged in other work, which usually involved an alternative

Table 7.4      Motivations for choice of teaching as a career,  
by sex, from 1978 data

Reason	Female	Male	$\chi^2$
Working with children	97 (91.5%)	51 (92.7%)	0.17
Standing of profession	62 (57.4%)	30 (54.6%)	0.16
Work conditions	99 (92.5%)	51 (92.7%)	0.00
Teacher as a family member	18 (16.7%)	16 (28.6%)	3.29
Exposure to a good teacher	42 (39.3%)	31 (55.4%)	4.25*
Nothing else appealed	44 (41.5%)	31 (56.4%)	3.35
* p < .05			

---

Table 7.5    Type of teacher training undertaken, by sex, from 1983 data

---

Subject area	Female	Male
Maths/science/commerce	18 (16.5%)	31 (55.4%)
English/history/library	27 (24.8%)	8 (14.3%)
Geography	4 ( 3.7%)	0 ( 0.0%)
Other languages	7 ( 6.4%)	0 ( 0.0%)
Art/music	14 (12.8%)	6 (10.7%)
Infants/primary	31 (27.5%)	10 (17.9%)
Special education	8 ( 7.4%)	1 ( 1.8%)
	<u>109 (100%)</u>	<u>56 (100%)</u>
Sex difference	$\chi^2=31.33$ , df=6, p <.001	

---



application of their university degree. The trend for more females to have resigned from teaching by 1983 (20% of females vs 9% of males) reflects the change in females' work patterns reported in Table 7.1.

In 1978, prior to commencing their teaching careers, subjects had been asked to estimate how they would perform at work, rating their answers on a visual analogue scale allowing for a range from 0 to 8). High scores (as reported in Table 7.7) reflect endorsement of the statement. There was a sex difference in imagined pleasure to be derived from work, with females anticipating it to be more pleasurable. This difference was reiterated in 1983 (but not in 1988) when females scored more highly on a different scale rating occupational importance and satisfaction (see Table 7.16).

#### Experience of illness and help-seeking behaviour

The rates for illness and help-seeking patterns (Table 7.8) include the three males who did not complete the DIS interview. In the help-seeking categories, some subjects sought help from more than one source and are included under each category. Subjects are counted only once in the total professional help category.

There was no significant sex difference in reporting significant illness, operations or accidents in males over the three five-year periods (pre-1978, 1978-83, 1983-88) but there is a trend for females to report more episodes in the last five-year period (1983-88) due to incidents related to child-rearing (e.g. miscarriage, caesarian section, termination of pregnancy).

Subjects were asked what professional help they had ever sought for depression. When asked in 1983, there were no sex differences in those who reported seeking any professional help for depression up to that time. In the group who had sought help, 19 females (17% of females) and 11 males (20% of males) had found the help 'moderately' or 'very' useful ( $\chi^2=2.02$ , ns). When asked in 1988 about any professional help up to that

Table 7.6 Teaching practices and location in 1979\* and 1983, by sex,  
from 1983 data

	Location in 1979		Location in 1983	
	Female	Male	Female	Male
State school system	77 (70.6%)	37 (66.1%)	61 (56.0%)	32 (57.1%)
Catholic schools	6 ( 5.5%)	5 ( 8.9%)	3 ( 2.7%)	4 ( 7.1%)
Independent schools	4 ( 3.7%)	4 ( 7.1%)	2 ( 1.8%)	5 ( 8.9%)
Migrant & special services	4 ( 3.7%)	2 ( 3.6%)	6 ( 5.5%)	3 ( 5.6%)
Tertiary institutions	2 ( 1.8%)	1 ( 1.8%)	2 ( 1.8%)	1 ( 1.8%)
Travel or further study	1 ( 0.9%)	1 ( 1.8%)	1 ( 0.9%)	0 ( 0.0%)
Resigned from teaching	2 ( 1.8%)	0 ( 0.0%)	22 (20.2%)	5 ( 8.9%)
Never taught, other work	13 (11.9%)	6 (10.7%)	12 (11.0%)	6 (10.7%)
	109 (100%)	56 (100%)	109 (100%)	56 (100%)
Significance	$\chi^2 = 2.99$ , df=7, ns		$\chi^2 = 9.46$ , df=7, ns	

\* 1979 was first year after graduation

Table 7.7 Prediction of work experience, by sex, from 1978 data

	Female	Male	t-test
Work will.....			
be a great pleasure for me	6.20 (1.15)	5.76 (1.22)	2.19*
be as difficult as could be imagined	3.42 (1.55)	3.71 (1.70)	-1.06
not be emotionally draining at all	3.95 (1.52)	4.11 (1.65)	-0.57
giving me feelings of doubt and low self confidence a lot	3.20 (1.36)	3.04 (1.55)	0.65
* p < .05			

time, 29 females (27% of females) and 11 males (21% of males) found such help to be 'moderately or 'very' useful ( $\chi^2 = 0.71$ , ns).

There were sex differences in rates of help-seeking from general practitioners for depression. Subjects were also asked to estimate the number of visits (for any health reason) they had made to a general practitioner over the previous twelve months. In 1983 (for 109 females and 56 males), females reported an average of 3.7 visits, males 4.9 visits (including one male who had numerous visits because of chronic renal failure). The number of visits for males fell to 2.3 if the three males who subsequently left the study are excluded (one of these being the man with renal failure). There is then a significant sex difference in number of visits ( $t = 2.95$ ,  $p < .005$ ), with females being more likely to visit.

When asked in 1988, concerning the preceding twelve months, for 108 females and 53 males, the females reported more visits, with a mean of 4.7 visits, and the males a mean of 2.1 visits ( $t = 2.89$ ,  $p < .005$ ). However, this comparison does not include the three males who did not fully complete the questionnaires because of chronic illness and their inclusion would have increased the rates for males.

Females were more likely to have sought help from friends both in 1978 and 1988, leading to significant sex differences. In 1983 males reported seeking out friends at about the same rate as females (and at a higher rate than in either 1978 or 1988) so that there was no sex difference in help-seeking from friends at that time. When these data are considered together with data for reported levels of satisfaction with perceived social support reported in Table 7.3, they indicate a trend for females to retain their friends as an important source of social support when depressed (Table 7.8) and in times of stress (Table 7.3) where at 1988, 85% of females rated friends as giving 'ideal' or 'close to ideal' support but only 76% of females rated their partners thus. On the other hand, data in Table 7.3 reports that males reported 'ideal' or 'close to ideal' social support in times of stress from friends at the same

Table 7.8 Experience of illness and help-seeking behaviour when depressed, by sex, from 1978, 1983 and 1988 data

Patterns of illness	Assessment	Females	Males	$\chi^2$
Chronic illness	Pre-1978	7 ( 6.4%)	3 ( 5.4%)	0.10
Chronic medication	Pre-1978	6 ( 5.5%)	3 ( 5.4%)	0.01
Illness, operation or injury in past 5 years	1978-83	26 (23.9%)	14 (25.0%)	0.01
	1983-88	36 (33.3%)	14 (25.0%)	2.63
<u>Help-seeking when depressed</u>				
Professional help, total	Pre-1978	16 (14.7%)	6 (10.7%)	0.29
	1978-83	30 (27.5%)	15 (26.8%)	0.01
	1983-88	43 (39.8%)	18 (32.1%)	1.04
From general practitioner	1978-83	19 (17.4%)	9 (16.1%)	0.06
	1983-88	25 (23.2%)	9 (17.0%)	0.81
From psychiatrist	1978-83	7 ( 6.4%)	7 (12.5%)	2.53
	1983-88	8 ( 7.4%)	9 (16.1%)	3.45
From counsellor	1978-83	15 (13.8%)	6 (10.7%)	0.14
	1983-88	11 (10.2%)	5 ( 8.9%)	0.02
From friends	Pre-1978	74 (67.9%)	27 (49.1%)	6.03*
	1978-83	92 (84.4%)	48 (85.7%)	0.05
	1983-88	85 (78.0%)	31 (59.6%)	7.21*
Medication "for nerves"	Pre-1978	12 (11.0%)	4 ( 7.1%)	0.55
	1978-83	14 (12.8%)	9 (16.1%)	0.72
	1983-88	26 (24.1%)	15 (26.8%)	0.15
				* p < .05

rate as females only in 1983 (86% of males vs 85% of females) but by 1988, only 58% of males rated their friends and 85% of males had rated their partners thus.

In summary, both sexes reported similar there were similar patterns in overall rates of illness and help-seeking from any professional at 1978, 1983 and 1988. By 1988, females were more likely to have consulted a general practitioner and males to have consulted a psychiatrist but these differences were not statistically significant.

Females were more likely to seek out their friends in normal circumstances and when depressed, but this difference did not reach statistical significance in 1983, where male and female help-seeking patterns were more similar in this regard. The finding that males tended to be more satisfied (than females) with levels of support from partners in times of stress in 1988 may reflect a sex difference in patterns of help-seeking. Males may perceive the most readily available support as satisfactory (i.e. friends in 1983, partner in 1988), while females more actively retain friendships as a valued social support, and there was, by 1988, a perception that friends offered more satisfactory social support (particularly in times of stress) than their partners.

#### Drug and alcohol use

Alcohol and drug usage was not estimated in 1978. Table 7.9 reports data from 1983 and 1988 assessments. Fewer subjects of both sexes reported drinking alcohol at 1988 than at 1983. There were no sex differences in intake of alcohol although there was a trend for males to have a higher weekly consumption (based on mean consumption for those who drank alcohol).

In 1983, there were no sex differences in numbers of subjects who were smoking. By 1988, many of the group had ceased smoking, but there were more of such men, leading to a sex difference.

Table 7.9 Drug and alcohol use, by sex, from 1983 and 1988 data

Drug usage	Assessment	Female	Male	Significance
Drink alcohol	Up to 1983	103 (94.5%)	51 (91.1%)	$\chi^2 = 0.26$
	1983-88	80 (74.1%)	32 (60.4%)	$\chi^2 = 3.15$
Mean weekly intake of alcohol (grams)	In 1983	52.1 (53.7%)	72.3 (108.5)	$t = -1.69$
	In 1988	56.0 (42.6%)	83.8 ( 88.3)	$t = -1.70$
Smoke cigarettes	Pre-1983	30 (27.5%)	9 (16.1%)	$\chi^2 = 1.23$
	1983-88	19 (17.6%)	2 ( 3.8%)	$\chi^2 = 4.67^*$
Have tried cannabis	Up to 1983	34 (31.2%)	24 (42.9%)	$\chi^2 = 2.21$
	Up to 1988	39 (36.1%)	27 (53.8%)	$\chi^2 = 2.38$
Have used narcotics	Up to 1983	4 ( 3.7%)	2 ( 3.6%)	$\chi^2 = 0.17$
	Up to 1988	3 ( 2.8%)	2 ( 3.6%)	$\chi^2 = 0.02$

\*  $p < .05$

Subjects were asked whether they had used cannabis either regularly or socially. There were no significant sex differences in rates of those who reported having smoked either at 1983 or 1988 assessments, although there was a trend for males to be more likely to have done so. Of those who said that they had smoked cannabis, use was generally experimental, being confined to a total of five occasions or less (ever). Only one male was currently smoking cannabis daily and admitted that he had a problem with his current level of usage. No member of the cohort was currently addicted to heroin although one female was using heroin recreationally and one male had been arrested for possession of narcotics and had only recently ceased participation in a methadone programme.

#### Coping styles when depressed

Subjects were asked which of a variety of coping styles were utilised during depressive episodes (see Table 7.10). Females were far more likely to report spending money on themselves and over-eat when depressed at 1983 and 1988 follow-ups. In 1983, males were more likely to report engaging in reckless behaviour but not in 1988, while females were more likely (in 1988 only) to report spending time with friends.

#### Experience of life events

In 1979, 101 subjects in the cohort had responded to a request to rate a series of positive and negative life event items (on a 0 to 8 scale) in terms of the predicted impact each event would have had on themselves. The method (as described in Chapter 6) was designed to generate individual and group mean predictive scores for each item which could be used for comparison with actual ratings of impact of specific life events when they occurred subsequently. The aim of the technique was to allow comparison of anticipated, and subsequent actual effects to examine for sex differences (e.g. whether either sex was more likely to perceive life events as more threatening) and individual differences (e.g. whether a depressed individual was likely to ascribe 'effort after meaning' to negative life events preceding the onset of a depressive episode). The same technique was used for both positive and negative life events. As 101 subjects responded in 1979, individuals



Table 7.10      Coping styles utilised when depressed, by sex,  
from 1983 and 1988 data

Coping style	Year	Female	Male	$\chi^2$
Be reckless	1983	15 (13.9%)	20 (36.4%)	10.92***
	1988	16 (14.8%)	10 (19.2%)	0.21
Eat more	1983	48 (44.4%)	11 (20.0%)	8.40**
	1988	56 (51.6%)	15 (28.8%)	6.62*
Drink more alcohol	1983	27 (25.0%)	17 (30.9%)	0.38
	1988	25 (23.2%)	14 (30.9%)	0.11
Spend money on self	1983	53 (48.6%)	11 (20.0%)	11.73**
	1988	54 (50.0%)	10 (19.2%)	12.59***
Become careless with contraception	1983	2 (1.9%)	0 ( 0.0%)	0.07
	1988	1 (0.9%)	1 ( 1.9%)	0.00
Busy oneself in work or hobby	1983	51 (47.2%)	22 (40.0%)	0.50
	1988	56 (51.9%)	22 (42.3%)	0.93
Socialise	1983	28 (25.9%)	11 (20.0%)	0.42
	1988	29 (26.9%)	13 (25.0%)	0.00
Do something to get mind off problem	1983	54 (50.0%)	27 (49.1%)	0.00
	1988	66 (61.1%)	33 (60.0%)	0.00
Shop-lift	1983	1 ( 0.9%)	0 ( 0.0%)	0.00
	1988	0 ( 0.0%)	0 ( 0.0%)	0.00
Spend time with friends	1983	50 (46.3%)	23 (41.8%)	0.14
	1988	67 (62.0%)	23 (44.2%)	4.10*
Ignore the problem	1983	22 (20.4%)	7 (12.7%)	0.98
	1988	33 (30.6%)	14 (26.9%)	0.08
Think through the problem	1983	68 (63.0%)	38 (69.1%)	0.36
	1988	84 (77.8%)	40 (76.9%)	0.04
Take care of physical appearance	1983	23 (21.3%)	6 (10.9%)	2.03
	1988	17 (15.7%)	5 ( 9.6%)	0.65
Develop a new skill	1983	9 ( 8.3%)	3 ( 5.6%)	0.12
	1988	13 (12.0%)	5 ( 9.6%)	0.03
Day dream	1983	40 (37.0%)	22 (40.0%)	0.04
	1988	52 (48.2%)	25 (48.1%)	0.00
* p < .05,    ** p < .01,    *** p < .001				

predictive ratings were only available for the subjects and group mean predictive scores for each item were derived from the scores of the same 101 subjects.

When life events for the previous twelve months were fully reviewed in 1983, the 165 subjects in the cohort confirmed whether each life event had occurred and then rated the effect of such events on themselves on a 0 to 8 scale for each item. Three life event effect impact scores were obtained - the actual weighting given at 1983 for events over the past twelve months, the predicted weighting given to those same events in 1979 (for the 101 subjects who had previously completed the predictive inventory) and the group mean predicted score for each item estimated from the 1979 data set.

The positive and negative total life event scores represent two combined scores for perceived impact of all the events that subjects had encountered over the previous twelve months. Data for 165 subjects reported in Table 7.11 shows slight but non-significant trends for females to report higher scores than males for both positive and negative life events. When the subjects' scores were compared with their own predictive score (n=101) and the group mean predicted scores, there were no significant differences. These findings imply that there is no sex effect in anticipatory perception of the effect of life events and that the sexes had similar views on the degree of distress elicited by both positive and negative life events.

In 1988, subjects (n=161) were given a simplified check list of life events over the previous twelve months. Table 7.12 reports data on subjects who affirmed the presence of the various life events. There were no sex differences in the likelihood of having experienced such life events estimated as likely to precipitate depressive episodes (e.g. break-up of intimate relationship, illness, death in family).

Table 7.11 Subjects' reported life events scores for twelve months up to 1983 (for 165 subjects), with subjects' own and group mean predicted life event scores from 1979 data (for 101 subjects), all analysed by sex

Life event scores	Year assessed	Female mean (SD)	Male mean (SD)	t-test
Total positive life events	1983	394 (219)	367 (217)	0.74
Subjects' predicted score	1979	360 (223)	358 (236)	0.05
Group predicted score	1979	284 (161)	274 (149)	0.67
Total negative life events	1983	91 (108)	67 (100)	1.45
Subjects' predicted score	1979	97 (132)	84 (106)	0.52
Group predicted score	1979	93 (118)	72 ( 75)	1.40

Table 7.12 Life events over the last twelve months (1987-88), by sex, from 1988 data

Life event	Female	Male	$\chi^2$
Serious illness or accident	11 (10.2%)	6 (11.3%)	0.02
Dismissal from job	6 ( 5.6%)	2 ( 1.9%)	0.11
Unemployed for over a month	17 (15.7%)	5 ( 9.4%)	1.20
Break-up of relationship	20 (18.5%)	10 (18.9%)	0.00
Death of close family member	25 (23.2%)	9 (17.0%)	0.81
Birth of a child	39 (36.1%)	18 (34.0%)	0.07
Change of marital status	18 (16.8%)	4 ( 7.6%)	2.57

Table 7.13 Self-reported duration of episodes of depression from 1978, 1983 and 1988 assessments. On each occasion duration is reported for the previous twelve months, analysed by sex

	1978		1983		1988	
	Female	Male	Female	Male	Female	Male
<u>Duration</u>						
Minutes <sup>1</sup>	13 (13%)	4 ( 8%)	12 (11%)	9 (16%)	7 ( 6%)	6 (11%)
Hours <sup>1</sup>	44 (23%)	19 (40%)	60 (55%)	29 (53%)	46 (43%)	24 (47%)
Days <sup>2</sup>	35 (35%)	22 (46%)	29 (27%)	14 (25%)	46 (43%)	18 (35%)
Weeks <sup>2</sup>	6 ( 6%)	2 ( 4%)	5 ( 4%)	1 ( 2%)	4 ( 4%)	2 ( 4%)
Months <sup>2</sup>	1 ( 1%)	1 ( 1%)	3 ( 3%)	2 ( 4%)	4 ( 4%)	1 ( 2%)
Trivial <sup>1</sup> vs sustained <sup>2</sup>	$\chi^2=1.42$ , ns		$\chi^2=0.05$ , ns		$\chi^2=0.85$ , ns	

### Self-reported number of depressive episodes

At all three assessments, subjects were asked to estimate the number of discrete episodes of depression they had experienced over the previous twelve months. They were given the following definition designed to describe depression ranging from 'normal' to 'clinical': "a significant lowering of mood with or without feelings of guilt, hopelessness and helplessness, or a drop in one's self-esteem".

At base-line assessment, in 1978, there were no sex differences in such reports, with both males and females reporting a mean of 8.6 episodes in the previous year ( $t=0.99$ , ns). In 1983, both sexes reported an increased number of episodes, females reporting a mean of 21.5 (SD 50.5) and males a mean of 22.3 (SD 55.8), with no sex differences ( $t=0.93$ , ns); in 1988, females reported a mean of 8.5 (SD 9.0) episodes and males, 8.4 (SD 11.6) episodes, with no sex difference ( $t=0.95$ , ns). The increase in the number of episodes in 1983 coincided with the early years of teaching, but the rate returned to 1978 levels by 1988, with no sex difference apparent on any occasion. The reported duration of these episodes is reported in Table 7.13.

When average duration of such depressive episodes was divided into categories of 'trivial' (duration of minutes to hours) or 'sustained' (duration of days to months), there were no significant sex differences at 1978, 1983 or 1988. It is of interest to note that at each assessment less than 10% of the sample reported average episodes lasting weeks or months.

### Precipitants to depression

In 1978, subjects were asked "to consider some common precipitants to depression" in terms of the likelihood that these might cause depression in themselves. Table 7.14 reports the numbers of subjects who judged the various precipitants as "frequently" resulting in their becoming depressed.

Table 7.14 Likely precipitants for depression, from 1978 data, by sex

"Possible precipitant to depression"	Females	Males	$\chi^2$
Too many life changes over a period	28 (26.2%)	11 (20.0%)	0.75
Failure to attain an important goal	56 (52.3%)	39 (70.9%)	5.05*
Break-up of an important relationship	75 (70.1%)	36 (65.5%)	0.34
Failure to live up to own standards	76 (71.0%)	35 (63.6%)	0.88
Being rejected or distanced	83 (77.6%)	38 (69.1%)	0.06
A fall-off in support to self-esteem	54 (50.5%)	20 (36.4%)	2.86
"Frequently results in depression"	Females	Males	$\chi^2$
Too many life changes over a period	7 ( 6.5%)	4 ( 7.3%)	0.02
Failure to attain an important goal	29 (27.1%)	3 ( 5.5%)	9.37**
Break-up of an important relationship	40 (38.1%)	13 (23.6%)	3.08
Failure to live up to own standards	37 (34.6%)	10 (18.2%)	4.70*
Being rejected or distanced	38 (35.5%)	14 (25.6%)	3.20
A fall-off in support to self-esteem	16 (15.0%)	2 ( 3.6%)	4.70*
* p <.05,    ** p <.01			

Table 7.15    Assessment of personality style of subjects at interview  
for 98 females and 51 males interviewed in person in 1983  
and 99 females and 45 males interviewed in person in 1988

Personality assessment	<u>Assessment year and rate</u>			
	1983/K.W.		1988/R.C.	
	Female	Male	Female	Male
Effective style	59 (60.2%)	32 (68.6%)	65 (65.7%)	33 (73.3%)
Slight impairment	32 (29.4%)	15 (26.8%)	30 (30.3%)	7 (15.6%)
Moderate impairment	6 ( 5.5%)	3 ( 5.4%)	4 ( 3.7%)	4 ( 7.6%)
Chronic impairment	1 ( 0.9%)	1 ( 1.8%)	0 ( 0.0%)	1 ( 1.9%)
Effective v the rest	$\chi^2=0.02$ , ns		$\chi^2=0.52$ , ns	
Effective/slight Impairment v the rest	$\chi^2=0.03$ , ns		$\chi^2=1.57$ , ns	

Four items (i.e. "too many life changes", "break-up of an important relationship", "being rejected") were rated similarly by both sexes. Females were more likely to rate "failure to achieve an important goal", "failure to live up to own standards" and "a fall-off in support to self-esteem", despite the fact that the mean number of reported episodes was the same (both sexes had reported a mean of 8.6 episodes over the previous twelve months). Thus despite similar depression rates, the sexes appeared to differ in terms of differential life stressors.

#### Perception of self as a "nervous person" and interviewer assessment of personality

In 1983 and 1988, subjects were asked whether they saw themselves "as a nervous person". In 1983, 37% of females (n=40) and 29% of males (n=16) rated themselves as "nervous"; there were no significant sex differences ( $\chi^2=2.63$ , ns); while 31% of females (n=33) and 19% of males (n=10) rated themselves thus in 1988, again with no sex differences ( $\chi^2=6.63$ , ns).

At the same two occasions, the interviewer made an assessment of personality based on the interview (see Table 7.15). This was done for 98 females and 51 males personally interviewed by the author in 1983 and for 99 females and 45 males personally interviewed by Ms. Curtain in 1988. Those with 'effective personality' were seen as functioning well in their personal relationships and work situation; 'slight impairment' implied some neurotic traits which did not interfere with their day to day functioning in either sphere; 'moderate impairment' implied neurotic traits which interfered with their day to day function either at work or in their interpersonal relationships, and 'chronic impairment' indicated that they were so impaired that they were unable to sustain work or were having major continuing interpersonal problems related to personality style.



Table 7.16 Mean scores on depression and personality scales by sex, from 1978, 1983 and 1988 data

Measure	Year	Female	Male	t-test
		Mean (SD)	Mean (SD)	
Self-esteem+	1978	1.6 ( 1.6)	1.5 ( 1.3)	0.63
	1983	1.1 ( 1.4)	0.7 ( 1.2)	1.73
	1988	1.0 ( 1.4)	0.8 ( 1.1)	0.61
Trait depression	1978	30.5 (11.4)	32.3 (11.1)	0.90
	1983	30.9 (11.2)	31.7 (11.7)	0.44
	1988	28.2 (11.4)	31.7 (14.0)	1.57
State depression	1978	57.2 ( 6.1)	56.7 ( 5.5)	0.49
	1983	55.5 ( 6.3)	54.6 ( 4.6)	0.92
	1988	56.3 ( 8.1)	54.2 ( 6.9)	1.96
Dependency scale	1978	53.9 ( 9.5)	50.4 ( 9.4)	2.24*
	1983	54.2 ( 9.0)	50.1 ( 8.3)	2.89*
	1988	54.7 ( 9.0)	51.7 ( 8.7)	2.04*
Neuroticism	1978	9.5 ( 4.7)	8.0 ( 4.6)	1.97*
	1983	9.5 ( 5.1)	7.3 ( 4.8)	2.57*
	1988	9.2 ( 5.3)	7.8 ( 5.3)	1.69
Dysfunctional attitudes	1983	78.4 (15.1)	78.4 (15.0)	0.00
Interpersonal sensitivity	1988	92.2 (12.6)	89.7 (10.4)	1.57
Bem Scale				
Masculinity	1983	4.54 ( .65)	4.74 ( .61)	1.85
	1988	4.66 ( .81)	4.71 ( .55)	0.40
Femininity	1983	4.83 ( .51)	4.65 ( .43)	2.16
	1988	4.85 ( .53)	4.45 ( .54)	4.70**
Social desirability	1983	5.24 ( .52)	5.26 ( .43)	0.32
	1988	5.36 ( .51)	5.28 ( .48)	0.95
Occupational - satisfaction	1983	89.1 (16.3)	81.8 (13.3)	2.79*
	1988	85.1 (17.3)	82.2 (17.3)	0.98
- importance	1983	95.0 (13.5)	89.4 (10.9)	2.60*
	1988	95.1 (12.6)	95.1 (11.3)	0.01

+ Higher scores reflect a lower self-esteem  
 \* p <.05, \*\* p <.01

Table 7.17 Mean scores on scales measuring perception of relationships, by sex, from 1978, 1983 and 1988 data

Measure	Year	Female	Male	t-test
PBI scores				
Maternal care	1978	26.0 (7.7)	26.5 (5.5)	-0.38
	1983	25.8 (9.2)	27.2 (6.2)	-1.15
	1988	26.1 (8.9)	26.8 (6.5)	-0.61
Paternal care	1978	21.9 (9.5)	21.6 (7.5)	0.19
	1983	22.1 (9.5)	20.9 (9.3)	0.73
	1988	22.1 (9.2)	20.0 (8.2)	1.47
Maternal overprotection	1978	15.1 (8.2)	14.0 (6.2)	0.91
	1983	14.1 (8.6)	12.8 (7.5)	1.05
	1988	14.2 (9.2)	13.0 (7.1)	0.94
Paternal overprotection	1978	13.8 (7.8)	12.5 (6.6)	1.07
	1983	12.5 (7.5)	11.3 (11.8)	0.62
	1988	12.0 (8.2)	11.8 (6.9)	0.12
IBM scores				
Care from partner	1983	30.8 (5.2)	30.2 (5.8)	0.61
	1988	29.7 (6.4)	29.0 (6.8)	0.54
Control by partner	1983	6.2 (6.3)	7.7 (8.2)	-1.48
	1988	6.3 (6.3)	8.2 (6.4)	-1.60
				* p <.05, ** p <.01

When the groups were divided as shown in Table 7.15, there were no sex differences in evaluation of personality on either occasion of interview. It should be noted that the groups interviewed on each occasion are largely comprised of the same subjects but are not identical as a few subjects responded by mail on each occasion.

#### Self-report questionnaires

The scores on self-report measures used in 1978, 1983 and 1988 are all reported in Table 7.16. There are no sex differences in mean scores on the state or trait depression scales in 1978, 1983 or 1988. Females scored significantly higher than males on the dependency measure at 1978, 1983 and 1988, and on the EPI neuroticism measure at 1978 and 1983, while this latter comparison failed to reach statistical significance in 1988. There were no sex differences in self-esteem scores at any assessment, but there was a significant change in scores over time. Both sexes rated an increase in self-esteem over time, with the greatest improvement over the first five years, from 1978 to 1983 (females:  $t=3.37$ ; males:  $t=3.49$ ; both  $p < .001$ ). There was a trend for self-esteem to have improved more in males than in females over the ten years. Females rated their principal work as both more satisfying and as more important on the occupational scale than did the males in 1983, but there were no differences in 1988. While there were trends for the females to score more highly (i.e. as more feminine) than males on the Bem femininity sex role scale (which was statistically significant at ten years). Males scored themselves more highly on the masculinity scale (i.e. as more masculine) but this was not statistically significant. There were no sex differences on the social desirability subscale.

Table 7.17 reports data from the PBI scales administered at 1978, 1983 and 1988. One hundred and sixty five subjects provided full PBI data for mothers and 160 subjects for fathers, as 3 females and 2 males had been separated from their fathers through death or divorce. There was no sex difference in any of the PBI scores on any occasion.

Table 7.18 continued

	Year	Female	Male	$\chi^2$
Experience of feeling...				
self-critical	1983	70 (64.2%)	31 (55.4%)	1.02
	1988	66 (61.1%)	25 (48.1%)	1.93
hopeless	1983	26 (23.9%)	6 (10.7%)	3.38
	1988	33 (30.6%)	9 (17.3%)	2.53
helpless	1983	21 (19.3%)	4 ( 7.1%)	3.12
	1988	18 (16.7%)	8 (15.4%)	0.00
lacking drive	1983	54 (49.5%)	29 (51.8%)	0.00
	1988	56 (51.9%)	33 (63.5%)	1.48
lacking motivation	1983	56 (51.4%)	28 (50.0%)	0.00
	1988	59 (54.6%)	31 (59.6%)	0.18
tearful	1983	58 (53.2%)	4 ( 7.1%)	32.05***
	1988	71 (65.7%)	4 ( 7.7%)	45.19***
tired	1983	67 (61.5%)	31 (55.4%)	0.43
	1988	77 (71.3%)	22 (42.3%)	11.31**
feels no pleasure	1983	31 (28.4%)	16 (28.6%)	0.00
	1988	25 (23.2%)	16 (30.8%)	0.71
suicidal	1983	6 ( 5.5%)	2 ( 3.6%)	0.03
	1988	10 ( 9.3%)	3 ( 5.8%)	0.20
brood over past	1983	45 (41.3%)	20 (35.7%)	0.33
	1988	28 (25.9%)	19 (36.5%)	1.43
decrease weight	1983	7 ( 6.4%)	5 ( 8.9%)	0.06
	1988	9 ( 8.3%)	4 ( 7.7%)	0.00
increase weight	1983	18 (16.5%)	2 ( 3.6%)	4.75*
	1988	20 (18.6%)	4 ( 7.7%)	2.43
argumentative	1983	37 (33.9%)	12 (21.4%)	2.32
	1988	33 (30.6%)	12 (23.1%)	0.64
irritable	1983	76 (69.7%)	27 (48.2%)	6.83*
	1988	71 (65.7%)	32 (61.5%)	0.12
Total number of symptoms	1983	10.9 (SD 5.2)	8.4 (SD 4.9)	t=3.06**
	1988	11.3 (SD 5.0)	9.2 (SD 5.0)	t=2.50*
* p < .05,    ** p < .01,    *** p < .001				

Completion of IBM scales required the subject to be involved in a committed intimate relationship. Scores were obtained from 88 females and 46 males at 1983 and 90 females and 42 males in 1988. There were no sex differences in any of the IBM scores on either occasion.

#### Self-report depressive symptom check-list

In both 1983 (n=165) and 1988 (n=161), subjects rated the number of symptoms that they "generally felt during periods of depression", without consideration of whether such episodes were clinically significant. Table 7.18 reports frequencies of these symptoms, which were quite stable over time. On both occasions, the most commonly noted symptoms were feeling 'down', 'flat', and 'depressed', (rather than 'sad' or 'despairing'), lacking drive and becoming irritable.

There were several sex differences, the most significant being females' increased reporting of tearfulness. In 1983, the females were more likely to report weight gain, feeling self-pity and pessimism, while in 1988, females were more likely to report sadness, tiredness and increased appetite. In 1983 and 1988 females generally reported more symptoms. The mean level of 8-10 symptoms appears high for a normal group but it should be noted that some of the items were descriptors of depression such as feeling 'sad', 'down' and 'depressed' so that most subjects rated several of these as being present.

#### Summary of significant sex differences

At base-line, in 1978, the significant sex differences were that males were more likely to have an undergraduate degree in science or mathematics, while females were more likely to have a degree in the humanities.

Females rated more highly on dependency and EPI neuroticism self-report scales and were more likely to have sought help from friends, when depressed. Males were more likely to view failure to live up to personal goals as a potential depressogenic stressor.

In 1983, at five-year follow-up, females again scored higher on the dependency and EPI neuroticism scales and rated more highly on the scale measuring items that were important to them and gave them satisfaction in terms of their working conditions.

In 1988, at ten-year follow-up, there continued to be a significant sex difference in terms of higher rating by females on the dependency scale, although differences on the EPI neuroticism scale just failed to reach statistical significance. There were no significant differences in job satisfaction. On the sex role inventory, there were no differences in terms of the masculinity or social desirability ratings, but females scored higher on the femininity subscale (a statistically significant finding in 1988).

At 1988, there was a significantly higher number of females engaged in home duties and part-time work. In 1978 and 1988, females were more likely to have sought help from friends and, in 1988, reported higher levels of perceived social support from friends both in normal circumstances and times of stress.

In both 1983 and 1988, there were few differences in coping styles when depressed. Females were more likely to eat more and spend money on themselves, both of these being self-consoling coping styles. Males were more likely to engage in reckless behaviour at 1983 only.

In 1983 and 1988, women reported a higher mean number of symptoms on a self-report symptom check list.

### Absence of sex differences

At the commencement of the study, there were no differences in the number of responders relative to the overall numbers in the group and no differences in the parental social class of the subjects. As the study proceeded, equal numbers went into teaching or into alternative careers, generally using their undergraduate qualifications. Equal numbers of both sexes married and had children at the same rate over time.

There were no sex differences in terms of drug and alcohol use, nor utilisation of professional help when depressed.

Throughout the study there was a lack of sex differences in self-reported numbers of depressive episodes. Both sexes reported the highest episodes during the 1978-83 period (which generally coincided with their first five years at work).

There were no sex differences in state and trait depression scores on self-report measures and rates of major depression, measured by the DIS. For the self-report self-esteem measure, there was a statistically significant improvement in self-esteem for the whole group, with a trend for greater improvement in self-esteem in males than females, but with no sex differences.

### Comment

The first aim of the study was achieved in that the group were initially homogeneous in terms of such psychosocial variables as age, marital and parental state. There were no sex differences in career aspirations, experience of life events and tendency to seek professional help for depression, which is against expectation but may be explained by the fact that members of both sexes are well-educated and have low levels of reported social deviance or drug and alcohol abuse, which in other groups has been ythought may be an alternative to help-seeking in males (Weissman & Klerman, 1977; Blazer et al, 1985).

There are some sex differences in scores on self-report measures and reported symptoms and coping styles when depressed. These are few in number and predictable in terms of sex-roles (e.g. females crying, tending to eat more and spend more money, males being reckless when depressed in their twenties). Indeed they are reassuringly present as these behaviours are expected sex-role-typed behaviours.

There has been some social role diversity over time (as hoped). While both sexes have been marrying and becoming parents at the same rate but over the ten year period, females were more likely to leave the full-time work-force to raise children and to work part-time.

These data provide the psychosocial context against which the reported rates of depressive disorders reported in Chapter 8 can be viewed.



## CHAPTER EIGHT

### INCIDENCE AND PREVALENCE DATA FOR DEPRESSIVE DISORDERS

#### CONTENTS

Introduction  
 Selection of diagnostic groupings  
 Use of the DIS  
 Examination of diagnostic groupings in terms of sex differences  
   Rates for depressive disorders from 1983 DIS assessment  
   Rates for depressive disorders from 1988 DIS assessment  
   Anxiety disorders from 1983 and 1988 DIS assessments  
 Relationship of anxiety to depressive disorders  
 Definition of caseness reconsidered  
 Summary

#### Introduction

The properties of the Diagnostic Interview Schedule or DIS (Robins et al, 1981), the case-finding instrument used in this study, were reviewed in Chapter 2. Some of the advantages of the DIS include its flexibility in defining diagnostic thresholds and the timing of episodes, so that DSM-III and RDC diagnoses can be obtained.

The depression diagnoses that can be generated from the DIS are DSM-III major depression (equivalent to RDC major depression, probable); DSM-III dysthymia (equivalent to RDC intermittent minor depression) and RDC minor episodic minor depression. These diagnoses will be reported in three diagnostic groupings: (i) major depression, i.e. DIS/DSM-III major depressive disorder, (ii) "DSM-III cases", i.e. DIS/DSM-III major depression and dysthymia, and (iii) "RDC cases", i.e. combined RDC categories for major depression (definite and probable), chronic intermittent minor and episodic minor depression (definite).

Cases are defined as subjects who have experienced episodes of depression that fall into these diagnostic groupings. There is an implied hierarchy in that, as major depression is included in each diagnostic grouping, any subject who has experienced an episode of major depression is automatically a member of all three groupings. Any subject with an episode of dysthymia (equivalent to RDC intermittent minor depression)

will be included as a 'DSM-III case' and 'RDC case', while subjects who have only experienced episodes of RDC minor depression will be classified as 'RDC cases' only.

'New cases' refer to subjects who have developed episodes of depression during the study. 'New case rates' report new cases as a proportion of the subjects at risk of developing a depressive disorder. Thus, subjects who had been diagnosed as cases at a previous assessment are no longer 'at risk' to become a new case at subsequent assessments, so that the denominator decreases over time. The formula for deriving 'new case rates' is shown at the foot of tables where new case rates are reported. 'Lifetime case rates' apply to any subject who has experienced a depressive episode prior to the designated assessment and are calculated as a proportion of the total number of subjects reviewed (which may be the cohort as a whole, or one or other sex).

#### Selection of diagnostic groupings

The first grouping contains only DSM-III major depression. This is the most widely reported category for depressive disorder in all the community studies using the DIS that were summarised in Chapters 2 and 3, with operationalised threshold criteria that should lead to high reliability.

The diagnoses constituting DSM-III affective disorders were outlined in Chapter 1. Of these, major depression, dysthymia and bipolar disorder constitute the categories reported for DSM-III affective disorders in most community studies using the DIS, and inclusion of this grouping would thus allow comparison of rates with other studies. In this study, the second grouping, termed 'DSM-III cases' is made up of the same DSM-III categories, however, there were no subjects reporting episodes of mania prior to 1983 and only one subject who reported an episode of mania in the 1983-88 period (thus attracting a diagnosis of bipolar disorder). As the subject had also reported episodes of major depression, she is included in the major depression category, rather than reporting a category of bipolar disorder for a single subject.

The reader is referred back to Table 1.1 for the diagnostic criteria for depressive disorders. In effect, DSM-III major depressive disorder (requiring 4/8 symptoms) is equivalent to RDC major depression (definite and probable). However, RDC allows for changes in minimum duration between 'probable' and 'definite' episodes of major depression (Table 1.1) and for inclusion of lifetime episodes with a requirement of one less symptom than for current episodes (in SADS-L). While this change in symptom threshold acknowledges the possibility that the number of symptoms reported can decrease with the passage of time, there is potential for confusion when current and past episode data for major depression are examined simultaneously. Thus, in this study, when rates for major depression are quoted, the DSM-III criteria for major depression will be used.

DSM-III or DSM-III-R dysthymia and RDC intermittent minor depression are very similar in concept, all requiring episodes with a minimum two years' duration but with subtle differences in wording to signify the amount of time required for the subject to be depressed during the episode (see Table 1.1). The third grouping, 'RDC cases', allows for the inclusion of RDC minor depression, for which DSM-III does not have an equivalent category. While the RDC minor depression category is not included in the DIS, it can be derived from DIS data by using the minimum duration of two weeks for episodes (required for 'definite' RDC minor depression) and setting a threshold of 2 of the possible 8 symptoms. The RDC impairment criteria are then imposed to give 'significant' episodes. However, as RDC minor depression has a diagnostic cut-off of 2 of a possible 16 symptoms, the threshold used in this study is higher.

Each of the diagnostic groupings is reviewed separately but subjects who had been allocated as RDC cases were still eligible to become DSM-III cases at a later date if the DSM-III case grouping (with a higher threshold for minimum symptoms or episode duration) was used instead. If a subject reported more than one type of

depression over the lifetime period, they were formally allocated to the diagnostic category with the highest threshold (i.e. RDC or DSM-III major depression took precedence over RDC intermittent minor depression or DSM-III dysthymia, which took precedence over RDC minor depression). Where subjects reported episodes of DSM-III major depression at an earlier assessment but were not found to reach criteria for DSM-III major depression at a subsequent assessment, they could still qualify for a diagnosis of RDC minor depression (and categorisation as RDC cases) if they had reached the minimum threshold of two symptoms and two weeks' duration required for RDC minor depression.

### Use of the DIS

As noted earlier, the DIS was given firstly (in 1983) by the author (at that time a psychiatrist with twelve years' clinical experience) and then five years later (in 1988) by a lay interviewer, Ms. Robyn Curtain (a graduate in psychology who was also a trained nurse).

At the 1983 review, estimates were made for RDC and DSM-III depressive categories for three periods: [a] lifetime episodes prior to base-line in 1978, [b] new cases 1978-1983, and [c] total lifetime episodes up to 1983 (i.e. [a] plus [b]).

At the 1988 review, estimates were made for RDC and DSM-III depressive categories (according to the same criteria used in 1983) for the following periods: [a] lifetime episodes up to 1978, [b] new cases 1978-1983, [c] lifetime episodes up to 1983, [d] new cases 1983-1988 and [e] total lifetime episodes to 1988 (i.e. [a] plus [b] plus [d]). This strategy allows assessment of whether a sex difference in rates of depression was present either at baseline or became apparent over the ten years of the study. Issues relating to consistency in reporting episodes from DIS-derived data over defined intervals will be addressed in the next chapter.

### Examination of diagnostic groupings in terms of sex differences

#### **Rates for depressive disorders from 1983 DIS assessment**

At the five-year review in 1983, when the mean age for males was 30 years and for females, 29 years, there were no statistically significant sex differences for any of the diagnostic groupings (major depression, DSM-III cases and RDC cases, as defined in the introduction to this chapter) and as reported in Table 8.1.

Examining the pre-1978 data first (calculated formally at the 1988 assessment), the lifetime rates for DSM-III major depression were 8% for females and 4% for males; for DSM cases, 11% for females and 4% for males; for RDC cases, 23% for females and 20% for males.

New case rates for the 1978-83 period for DSM-III major depression were 12% for females and 17% for males; for DSM-III cases, 14% for females and 20% for males; for RDC cases, 27% for both females and males.

Reporting lifetime rates to 1983, for DSM-III major depression, the rates are 19% for females and 20% for males; for DSM-III cases, the rates are 24% for females and 23% for males and for RDC cases, 44% for females and 41% for males. There were no subjects with a diagnosis of bipolar disorder at this stage.

#### **Rates for depressive disorders from 1988 DIS assessment**

In 1988, 161 of the 164 subjects contacted consented to DIS assessment. The second interviewer obtained data for all episodes to 1988 and thus made an independent assessment of each time period (i.e. lifetime rates to 1978 and 1978-83) covered by the author, allowing inter-rater consistency to be examined. Table 8.2 reports data derived from the 1988 DIS assessment.

Table 8.1 Numbers (rates) for new cases and lifetime cases for  
DSM-III major depressive disorder, DSM-III cases and RDC cases  
derived from 1983 DIS interview administered by the author

Interview in 1983	Female (n=109)	Male (n=56)	Total (n=165)	$\chi^2$
<u>DSM-III major depressive disorder (MDD)</u>				
Lifetime episodes to 1978	9 ( 8.3%)	2 ( 3.6%)	11 ( 6.9%)	0.66
New cases 1978-83	12 (12.0%)	9 (16.1%)	21 (12.7%)	0.46
Lifetime episodes to 1983	21 (19.3%)	11 (19.6%)	32 (19.4%)	0.02
<u>DSM-III Cases (MDD and dysthymia)</u>				
Lifetime episodes to 1978	12 (11.0%)	2 ( 3.6%)	14 ( 8.5%)	1.76
New cases from 1978-83	14 (14.4%)	11 (21.6%)	25 (16.6%)	0.55
Lifetime rates to 1983	26 (23.8%)	13 (23.2%)	39 (23.6%)	0.01
<u>RDC Cases (major, intermittent &amp; episodic minor depression)</u>				
Lifetime episodes to 1978	25 (22.9%)	11 (19.6%)	36 (21.8%)	0.08
New cases 1978-83	23 (27.4%)	12 (26.7%)	35 (27.1%)	0.02
Lifetime episodes to 1983	48 (43.5%)	23 (41.1%)	71 (43.0%)	0.13
* p <.05				
Lifetime case rates = $\frac{\text{number of subjects with episode to that time}}{\text{number of subjects reviewed}} \times 100\%$				
New case rates = $\frac{\text{number of new cases}}{\text{number of subjects reviewed - previous cases}} \times 100\%$				

Table 8.2 Numbers (rates) for new cases and lifetime episodes for  
DSM-III major depressive disorder, DSM-III cases  
and RDC cases derived from 1988 DIS interview administered  
by interviewer R.C.

Interview in 1988	Females (n=108)	Males (n=53)	Total (n=161)	$\chi^2$
<u>DSM-III major depressive disorder (MDD)</u>				
Lifetime episodes to 1978	8 ( 7.4%)	4 ( 7.6%)	12 ( 7.6%)	0.08
New cases 1978-83	10 ( 9.3%)	4 ( 7.6%)	14 ( 8.7%)	0.00
Lifetime episodes to 1983	18 (16.6%)	8 (15.1%)	26 (16.1%)	0.07
New cases 1983-88	10 (11.1%)	1 ( 2.2%)	11 ( 7.9%)	2.09
Lifetime episodes to 1988	28 (25.9%)	9 (17.0%)	37 (23.0%)	1.61
<u>DSM-III Cases (MDD and dysthymia)</u>				
Lifetime episodes to 1978	9 (21.3%)	4 ( 7.6%)	13 ( 8.1%)	0.00
New cases 1978-83	14 (14.1%)	4 ( 8.2%)	18 (12.2%)	0.61
Lifetime episodes to 1983	23 (21.3%)	8 (15.1%)	31 (19.3%)	0.88
New cases 1983-88	10 (11.8%)	2 ( 4.4%)	12 ( 9.2%)	1.56
Lifetime episodes to 1988	33 (30.6%)	10 (18.9%)	43 (26.7%)	2.48
<u>RDC Cases (major, intermittent &amp; episodic minor depression)</u>				
Lifetime episodes to 1978	20 (18.5%)	8 (15.1%)	28 (17.4%)	0.59
New cases 1978-83	27 (30.7%)	6 (13.3%)	33 (24.8%)	4.80*
Lifetime episodes to 1983	47 (43.5%)	14 (26.4%)	61 (37.9%)	4.42*
New cases 1983-88	14 (13.0%)	5 ( 9.4%)	19 (11.8%)	1.60
Lifetime episodes to 1988	61 (56.5%)	19 (35.8%)	80 (49.7%)	6.05*

\* p < .05

Lifetime case rates =  $\frac{\text{number of subjects with episode to that time}}{\text{number of subjects reviewed}} \times 100\%$

New case rates =  $\frac{\text{number of new cases}}{\text{number of subjects reviewed - previous cases}} \times 100\%$

Examining the pre-1978 data first, the lifetime rates for DSM-III major depression were 7% for females and 8% for males; for DSM-III cases, 8% for both females and males; for RDC cases, 19% for females and 15% for males, with no significant sex differences.

New case rates for the 1978-83 period for DSM-III major depression were 10% for females and 8% for males; for DSM-III cases, 14% for females and 8% for males; for RDC cases, 31% for females and 13% for males, with a statistically significant sex difference for the last grouping.

Reporting lifetime rates to 1983, for DSM-III major depression, the rates are 17% for females and 15% for males; for DSM-III cases, the rates are 21% for females and 15% for males and for RDC cases, 44% for females and 26% for males, with a statistically significant sex difference for the last grouping. There were no subjects with a diagnosis of bipolar disorder at this stage.

There had been no sex differences in rates for any diagnostic groupings to 1983 when data were collected by K.W. at 1983. From the 1988 data, there were statistically significant sex differences in new case rates (1978-83) and lifetime episode rates to 1983. Comparisons of rates reported from the two independent DIS assessments for the same interval and the implications of the sex differences in rates obtained at the 1988 assessment will be discussed in the following chapter.

Between 1983 and 1988 (see Table 8.2), new case rates for DSM-III major depression category were 11% for females and 2% for males; for DSM-III cases, 12% for females and 4% for males; for RDC cases, 23% for females and 13% for males. One of the females in the major depression category also reported an episode of mania.



Lifetime prevalence rates up to 1988 were 26% for females and 17% for males for DSM-III major depression; 31% for females and 19% for males for DSM-III cases; 57% for females and 36% for males for RDC cases, with a statistically significant sex difference for the last analysis.

While there are no sex differences in numbers of "new cases 1983-88" for any of the diagnostic groupings, there was a trend for females to report more episodes. At 1988, there was a non-significant trend for a female preponderance in lifetime rates for DSM-III major depression and DSM-III cases to 1988 and a statistically significant sex difference for RDC cases to 1988.

#### **Anxiety disorders generated from 1983 and 1988 DIS assessments**

Anxiety disorder diagnoses were also generated at the 1983 and 1988 DIS interviews. Although the DIS Eating Disorders section was not administered, one subject volunteered information concerning four years' treatment for anorexia nervosa (and did not have a diagnosis of any depressive disorder). Two other females who had been categorised as having episodes of RDC minor depression volunteered episodes of bulimia.

##### **i) Simple phobia**

In 1983, nine females and 3 males reported a simple phobia ( $\chi^2=0.13$ , ns). Of the 12 subjects, 9 had been categorised as RDC cases and 6 as DSM-III cases for depression at the same 1983 assessment. In 1988, ten females and one male reported a simple phobia ( $\chi^2=1.99$ , ns), in all cases the phobias were reported as present for at least five years. Five of these were the same subjects as at 1983 assessment. Six of the eleven had been classified as DSM-III cases for depression at 1983, and 10 as RDC cases at 1983.

ii) Social phobia

In 1983, 4 females and 3 males had reported a social phobia ( $\chi^2=0.01$ , ns), although for 2 of the females and 2 of the males, the symptoms were concurrent with episodes of major depression. Of the other three subjects, one had not reported any depressive episodes, one had also reported an episode of dysthymia (i.e. was a DSM-III case), one an episode of minor depression (i.e. an RDC case). In 1988, four females and one male were categorised as having a social phobia ( $\chi^2=0.02$ , ns). Two of the females had been similarly classified at 1983 assessment, while one female and one male had reported the onset of social phobia since 1983, but had both previously reported episodes of minor depression. All four had also been classified as RDC cases (and 3/5 as DSM-III cases) at 1983 assessment.

iii) Generalised anxiety disorder

Four females and two males had reported episodes of generalised anxiety disorder at 1983 assessment ( $\chi^2=0.17$ , ns), but in only 2 females and 1 male were these episodes independent of a depressive episode. All 6 subjects had also been allocated as RDC cases (and 5/6 as DSM-III cases) at 1983. At 1988, 3 females and 2 males were diagnosed as having generalised anxiety disorder ( $\chi^2=0.02$ , ns), of whom one female had previously reported such episodes concurrent with an episode of major depression. The four other subjects reported new episodes for the 1983-88 period but all had previously been classified as RDC cases (and 3/4 as DSM-III cases) at 1983.

iv) Panic disorder and agoraphobia

At 1983 assessment, 8 females and 4 males reported episodes of panic disorder ( $\chi^2=0.07$ , ns), of whom 2 females and 3 males also reported agoraphobia ( $\chi^2=0.59$ , ns). All of the subjects were RDC cases (and 7/12 were DSM-III cases) for depression. For 7 of the twelve reporting panic disorder, and for 2 of the five with agoraphobia, their symptoms were concurrent with episodes of depression.

Nine females and five males subjects reported episodes of panic disorder at 1988 assessment ( $\chi^2=0.04$ , ns). Seven of these (5 females and 2 males) had reported episodes starting prior to 1978, and all had reported such episodes at 1983. Thirteen of the fourteen had been classified as RDC cases for depression, and eight of the fourteen subjects had been classified as DSM-III cases at 1983.

Eight subjects (4 female and 4 male) reported onset of panic disorder since 1983 ( $\chi^2=0.45$ , ns). Of the eight subjects, four (2 female, 2 male) had previously reported episodes of major depression and then failed to report these earlier episodes at the 1988 DIS assessment. Instead, they reported the onset of episodes of panic disorder independent of major depressive episodes, although one female also reported 2 subsequent episodes of major depression (both post-partum). Of the other four subjects, three (1 female and 2 males) had previously reported episodes of minor depression at 1983 (but the 2 males failed to report these at 1988 assessment). When the three subjects who had reported earlier episodes of minor depression at 1983 were assessed at 1988, the female had experienced an episode of major depression following the birth of her first child and one male had experienced episodes of dysthymia and minor depression since 1983, the other also was diagnosed as having generalised anxiety disorder and minor depression. The one female who was not an RDC case for depression at 1983 had, at 1983, reported two simple phobias and two episodes of 'post-partum depression' which did not reach criteria for a depressive disorder at 1983. These findings demonstrate that the subjects developing panic disorder often had attracted multiple diagnoses over time and also raise the possibility that recall of previous episodes of depression may be impaired by the severity of subsequent symptoms such as panic disorder.

#### Relationship of anxiety disorders to depressive disorders

At 1983, a diagnosis of anxiety disorder was made for 22 subjects, 16 (15%) of the females and 6 (11%) of the males, with no sex differences ( $\chi^2=0.05$ , ns). Both

anxiety and depressive disorders were reported in seventeen of the 22 subjects. For 12 of the 17, the anxiety and depression diagnoses were concurrent. For the other five, the anxiety disorder preceded the onset of depression. All of the diagnoses of panic disorder/agoraphobia and generalised anxiety disorder occurred in subjects who also reported episodes of depression (i.e. the subjects had been categorised as being at least RDC cases) and the episodes were often, but not necessarily concurrent. Multiple concurrent diagnoses for anxiety disorders were also often made.

At 1988, there were four subjects (2 female, 2 male) who reported simple phobia, with no other diagnosis of anxiety or depression. When considering anxiety disorders other than simple phobia, there had been 19 subjects who had reported anxiety disorders other than simple phobia (the least reliably reported category) at any time up to 1983. Of those 19, eighteen had been classified as RDC cases (11/19 as DSM-III cases) at 1983.

For the nine subjects (6 females and 3 males) reporting first onset of anxiety disorders (there were no new onsets for simple phobia) after 1983, all had been classified as RDC cases and 5/9 as DSM-III cases at 1983 assessment. Four of those (2 females and 2 males) diagnosed as having panic disorder (one with agoraphobia) since 1983, failed at 1988 to report episodes of depression reported at the 1983 assessment which may have been influenced by the effect of subsequent onset of panic disorder.

Thus, there is an overlap between experience of anxiety and depression. If one was to report lifetime prevalence of combined anxiety disorders and combined RDC depression categories, the rates would be virtually the same as for RDC cases alone, as all but one of those reporting anxiety disorders (other than simple phobia) had already been subsumed as an RDC case of depression. Using DSM-III case criteria, 58% of the

group had attracted such a diagnosis, so that there seems to be more separation of experience of anxiety and depression when the threshold for depressive disorder is higher. There is still considerable overlap and co-occurrence of anxiety and depressive disorders and the issue of co-morbidity will be discussed further in Chapter 11.

### Definition of caseness reconsidered

In an earlier chapter examining concepts of caseness, the conclusion was reached that caseness definitions should be relevant to the question being considered (see Chapter 1). For this study of young adults, depressive caseness is defined in terms of (a) sustained depressive symptomatology, (b) help-seeking because of depression, and (c) impaired psychosocial functioning. A second definition of caseness will now make use of data available from more than one source (i.e. multiple points of assessment) and the addition of prolonged absence from work (where due to depression) to the RDC impairment criteria.

By this new definition, 'defined cases' includes subjects who had reported the same depressive episodes on at least two assessments. At the first assessment the minimum requirements were a report of: (i) for data from 1978 assessment, depressive episodes lasting weeks or months for which help was sought or medication taken on more than on a single occasion, or where subjects had attempted suicide concurrent with a reported depressive episode, or, (ii) for data from 1983 assessment, at least one episode of DSM-III major depression (with RDC impairment criteria fulfilled, or an episode of minor depression (with RDC impairment criteria fulfilled) and a minimum duration of 12 weeks, or an episode of minor depression (with RDC impairment criteria fulfilled) where subjects had taken long periods of leave (at least 3 months) from work as a result of experiencing such depressive episodes, and there was a later categorisation of either major depression for the same episode (with RDC impairment criteria) or minor depression (with RDC impairment criteria). Episodes were disallowed if help-seeking comprised a sole consultation with a general

practitioner in a crisis situation (e.g. break-up of intimate relationship) without any ongoing counselling, medication or significant impact on the subjects' life.

(iii) For data from the 1988 assessment covering new episodes from 1983-88 period, it was not possible to verify the episodes on more than one assessment occasion. Here all episodes of major depression were included, as well as all episodes of minor depression lasting longer than 12 weeks (which includes dysthymia) or leading to an absence of at least three months from work. RDC help-seeking criteria were again imposed as detailed above (see previous paragraph).

#### Rates for 'defined cases'

By 1983, there were 52 'defined cases' within the total sample of 165 (35/109 or 32.1% of females and 17/56 or 30.4% of males), with no sex differences ( $\chi^2=0.05$ , ns). All those who were cases at 1983 were automatically included at 1988, with there being 54 cases (37/109 or 33.9% of females and 17/56 or 30.4% of males) of the total of 165 at 1988. However, four subjects withdrew from the 1988 assessment, of whom three had been defined as cases, so that at 1988, 51 subjects (36/108 or 33.3% of females and 15/53 or 28.3% of males) were reported as 'defined cases' of the total of 161 subjects, again with no sex differences ( $\chi^2=0.42$ , ns). There were only two new cases (both females) for the 1983-88 period using this method.

The 'defined case' category was the only one which took into account data reported at 1978, which was then corroborated by data from 1983 and 1988 assessments. The negligible change in rates for 'defined cases' from 1983 to 1988 also demonstrates that some subjects who are reported to be 'new cases' in RDC and DSM-III systems had reported episodes of depression, help-seeking or impairment due to depression at previous assessments. For the 161 subjects examined on both occasions, the lifetime rates were 33% for females and 28% for males using this definition, with a non-significant sex difference.

The lifetime rates derived by the 'defined case' method lay between the DSM-III and RDC case rates. The higher rates prior to 1978 and lower 'new case' rates for 1983-88 period point to the possibility that subjects who present as new cases have had previous episodes that have been forgotten and as a corollary, those subjects who present early in life as cases are picked up by a variety of case-finding techniques used at different occasions over their life-span.

### Summary

There were no significant sex differences for any diagnostic categories for data to 1983 derived at the 1983 DIS assessment.

There were no significant sex differences for DSM-III major depression, with or without dysthymia, for any interval to 1988, from derived at the 1988 DIS interview. However, there is a trend towards female preponderance in new cases in the 1983-88 interval. There are statistically significant sex differences for 'new RDC cases' for 1978-83 interval and for lifetime prevalence using 1988 data only.

There are statistically significant sex differences in rates for combined RDC categories when data from 1988 assessment are used that were not apparent from 1983 assessment data.

A continuity between anxiety and depressive disorders was demonstrated in that almost all of those subjects who reported anxiety disorders were also diagnosed as RDC cases for depression and about 60% were also diagnosed as DSM-III cases of depression. In most cases, the diagnosis of depression preceded or was concurrent with the diagnosis of anxiety disorder. Virtually all subjects who reported anxiety disorders (other than simple phobia) up to 1988 had already been identified as RDC cases of depression at 1983.

A further definition of lifetime caseness was described. Episodes that did not fulfill criteria for major depression at each DIS assessment were allowed if they fulfilled criteria for minor depression (with a minimum duration of 12 weeks at the first report or caused the subject to take extended leave from work) or further operationalised impairment criteria that were applied to data derived from semi-structured interview or self-report data. This method of case definition led to lifetime rates that lay between those for DSM-III cases and RDC cases, with no significant sex differences.

#### Note

The finding that there were sex differences in rates at 1988 assessment that were not apparent at 1983 assessment will be further considered in the following chapter which examines issues concerning reliability of data.



## CHAPTER NINE

### RELIABILITY OF DATA

#### CONTENTS

Introduction  
 Data from self-report measures  
 Data from DIS-derived diagnostic groupings  
     Data from 1978 assessment  
     Collaborative witness reports at 1983 assessment  
     Inter-rater reliability between the two independent  
     (1983 and 1988) raters  
     Comparison of rates for diagnostic groupings for  
     periods prior to 1983, from 1983 and 1988 DIS data  
     Comparison of reported numbers of symptoms and duration  
     of episodes to 1983, from 1983 and 1988 DIS data  
     Consideration of various methods of estimating lifetime  
     case rates using 1983 and 1988 DIS-generated data  
     Further examination of collaborative witness reports at  
     1983 interview to examine disparities in 1983 and 1988  
     DIS data  
 Data concerning recourse to help-seeking  
 General discussion of reliability of DIS-derived data  
 Summary

#### Introduction

This chapter describes a series of studies examining aspects of the consistency and reliability over time of the self-report measures and DIS-derived diagnostic groupings used in the study. Some of the self-report data were collected at three assessment points at five-year intervals (1978, 1983 and 1988), while other data (including the DIS-derived diagnostic categories) were collected at the 1983 and 1988 assessment points only. The DIS-derived data for the period prior to 1983 will be examined in terms of long-term reliability and the use of a physician and as against a lay interviewer.

#### Data from self-report measures

Table 9.1 reports the means for the self-report measures which were used at 1978, 1983 and 1988 and their levels of consistency. For the IBM scale, a separate analysis was done for subjects who rated the same partner on both occasions.

Table 9.1 Mean scores and correlation coefficients for self-report measures used at 1978, 1983 and 1988

	Mean score			Correlation coefficients		
	1978	1983	1988	A with B	B with C	A with C
	(A)	(B)	(C)			
<u>Personality and trait scales</u>						
Neuroticism	9.0	8.7	8.7	.54	.68	.50
Self-esteem <sup>#</sup>	1.6	1.0	.9	.43	.61	.48
Dependency	52.6	52.8	53.7	.64	.64	.55
Trait Depression	31.3	31.2	29.4	.63	.65	.46
<u>Sex Role Inventory</u>						
Masculinity	N/A	4.6	4.7	N/A	.56	N/A
Femininity	N/A	4.7	4.8	N/A	.57	N/A
Social desirability	N/A	5.2	5.3	N/A	.57	N/A
<u>State depression</u>	57.0	55.2	55.3	.25	.23	.17
<u>PBI scales</u>						
Maternal care	26.3	26.2	26.3	.72	.82	.63
Maternal protection	14.8	13.8	13.8	.74	.76	.68
Paternal care	21.9	21.7	21.4	.80	.82	.72
Paternal protection	13.0	12.1	11.9	.69	.67	.56
<u>IBM scales</u>						
Intimate care	N/A	30.6	29.5	N/A	.52	N/A
Intimate control	N/A	6.7	6.9	N/A	.37	N/A
Intimate care (same partner)*	N/A	31.5	30.2	N/A	.48	N/A
Intimate control	N/A	6.5	7.0	N/A	.40	N/A

N/A = Not assessed

\* Those nominating the same intimate in 1983 and 1988 (n=96)

# Higher scores indicate lower self-esteem

The self-esteem, neuroticism, trait depression, dependency and sex-role scales are described in Chapters 9 and 10 and were included to measure trait personality dimensions. Moderate to high correlations between scores over time would be expected if these measures assess constructs that are stable and unaffected by current affective state. Conversely, one would predict that state depression scores would be poorly correlated over a ten-year period. Of the two measures that quantified inter-personal constructs, one would predict that the PBI sub-scales (measuring perceptions of the first sixteen years of the subjects' lives) would be more highly correlated on successive assessments over time than the IBM sub-scales (measuring perceptions of the current relationship with intimate partner) when the PBI strategy measures a 'constant' while the IBM assesses recent and current characteristics that may have indeed changed over the interval.

Mean scale scores for most personality measures were generally stable, apart from self-esteem scores, which showed a statistically significant improvement from 1978 to 1983 ( $t=4.80$ ,  $p < .001$ ) and from 1978 to 1988 ( $t=5.82$ ,  $p < .001$ ) but no change from 1983 to 1988 ( $t=0.87$ , ns). In a similar but less distinct fashion, depression scores decreased from 1978 to 1988, both on the trait ( $t=2.03$ ,  $p < .05$ ) and state ( $t=2.21$ ,  $p < .01$ ) measures but trait depression scores remained moderately well correlated with each other over the ten years. Scores for state depression were poorly correlated (Table 9.1). Sex-role inventory scores were unchanged over the 1983 to 1988 period for the femininity ( $t=1.37$ , ns) and masculinity ( $t=-1.41$ , ns) sub-scales. There was a significant difference in the social desirability scores ( $t=-2.43$ ,  $p < .05$ ) but this sub-scale is included in the Bem Sex Role Inventory as a distractor rather than as a measure of sex-role characteristics and may be influenced by state effects.

Next, the measures of interpersonal constructs are examined. When data collected at 1978 and 1988 are compared, no significant changes are evident in maternal

care ( $t=-0.02$ , ns), and over-protection scores ( $t=0.58$ , ns), nor paternal care scores ( $t=1.56$ , ns). All scores show moderate to high correlations (Table 9.1). Paternal overprotection scores did show a significant change over the ratings at 1978 and 1988 ( $t=2.12$ ,  $p < .05$ ) and were less highly correlated, although correlations still ranged from 0.56 to 0.69.

When data collected at 1983 and 1988 are compared, the IBM care scores showed a statistically significant change ( $t=2.68$ ,  $p < .01$ ) for the whole group and for those subjects who had the same partner on both occasions ( $t=2.86$ ,  $p < .01$ ), while there was no change in IBM control scores ( $t=-1.06$ , ns) for the group or those who had the same partner on both occasions ( $t=-0.89$ , ns). Although there are statistically significant differences in the PBI paternal overprotection and IBM care scales, the mean scores for both scales are quite similar. The differences may be due to small standard deviations and do not appear to be clinically significant.

#### Discussion of reliability of self-report data

Scores are generally more consistent in the interval 1983-1988, while the data for the 10-year interval (1978-1988) show slightly less consistency, presumably reflecting changes in attitudes in the first five years of the study. The four measures quantifying aspects of personality on three occasions return mean correlations of 0.56 (1978-1983), 0.64 (1983-1988) and 0.50 (1978-1988) respectively. The sex-role inventory scores are only reported in 1983 and 1988 and fall within the same range (Table 9.1).

The mean correlation coefficients for PBI scores are 0.74 (1978-1983), 0.77 (1983-1988) and 0.65 (1978-1988) showing an impressive test-retest consistency in PBI scores for all three intervals and superior to each individual personality test. The correlation coefficients for IBM scores (both for the whole sample and for the sub-group scoring the same partner) are moderate from 1983 to 1988 and clearly less stable

than PBI scores. As noted, the PBI and IBM scales are similar in design, there is a major difference in that PBI scales quantify perception of past relationships (from first 16 years of subject's life) which should only change if the subjects' perception of those relationships undergoes some change after the age of 16. The IBM scales quantify perception of a current relationship which is still evolving (mean duration of marriages was 7 to 8 years in 1988) and in a process of evolution as they were also becoming parents. Ideally, the measure will be more stable if these factors are taken into account. The stability of PBI scores builds to the reliability and validity of the measure in that subjects seem to report relatively constantly over a long period when many of them had been depressed and experienced current life changes (starting to teach, becoming parents) which may have been expected to exert a state effect. Finally, the test-retest reliability of the state depression measure is low, as one would anticipate. The stability of self-report data over time is in the predicted direction (with trait measures being more constant than state measures). The pattern of results suggests reasonably reliable reporting on self-report measures by the cohort.

### **Data for DIS-derived diagnostic groupings**

The selection of diagnostic groupings has been discussed in the previous chapter, where rates for DIS-derived diagnoses gathered at 1983 and 1988 assessments were reported and examined in terms of sex differences. As the test-retest reliability studies for structured interview schedules (including the DIS) noted in Chapter 1 were for much shorter time periods than those in this study, it is important to review the long-term reliability of the present data.

There are several possible sources of error: (i) changes due to differences in recall by subjects over time, either leading to forgetfulness or changes in attribution (so that episodes are considered as depression on one occasion, but distress on another), (ii) differences in style of interview or interpretation of the DIS-derived data due to administration by two different raters.

The choice of the three different diagnostic groupings (identified in Chapter 8) represented different definitions and diagnostic thresholds for caseness and were examined to determine whether sex differences were more apparent at one threshold of caseness than another. Subjects who reported differing numbers of symptoms at the two DIS assessments may then have been retained as cases for the grouping with the lowest threshold (i.e. RDC cases) so that lifetime caseness allowing formal "agreement" for both occasions, despite reported unreliability of lifetime caseness for a grouping with a higher threshold (e.g. DSM-III major depression only) which would have given a formal "disagreement" for the two interview occasions.

i) Data from 1978 assessment

At base-line in 1978, the DIS was not used and there was no attempt to corroborate the self-report information gathered.

ii) Collaborative witness reports at 1983 assessment

At the five-year follow-up, after the personal interview, each subject was asked to nominate another person who had known the subject well for a number of years and who might be able to provide an independent account of the subject's depressive experience to the time of the 1983 assessment. Consent and a completed corroborative interview were obtained for 133 of the 150 (88.7%) subjects interviewed in person, and therefore 80.6% of the whole cohort. When possible, the nominated witness was interviewed immediately to avoid discussion and "priming" by the subjects, and this occurred for 56 of the 133 (42.1%) corroborative interviews. If the person nominated was not available, the subject was asked to inform that witness that there would be telephone contact in the next few days but not to describe the exact nature of the information to be sought. This procedure was adopted for the remaining 77 subjects. On a number of occasions, more than one witness was sought to clarify details. The method and form of questions put was described in detail in Chapter 6.

Where subjects had reported episodes of major depression, 98% of witnesses volunteered accounts of episodes consistent with those reported by the subjects. Of the 32 subjects reporting an episode, 26 volunteered contactable witnesses, with 23 of the 26 agreeing with the subjects' reports as to occurrence and impact (impairment). In the three instances of dissonant views, two informants confirmed the timing but considered the episode less disabling than judged by the subject, while one witness failed to confirm severity or timing (Table 9.2).

For the 54 subjects reporting RDC minor or intermittent minor depression, 49 witnesses were contacted and 41 of these volunteered accounts that agreed with their subject's view about timing and impact of any episodes reported. Three witnesses judged episodes as having occurred which had not been volunteered by subjects. For five subjects reporting episodes, two witnesses had not noted the episodes and an additional three agreed with the timing but did not judge the episode as having had a significant impact on the subject's life (Table 9.2).

#### Discussion of collaborative witness data

These data show high reliability for presence and timing of depressive episodes but no attempt was made to distinguish between major and minor depressive episodes (i.e. by asking for frequency of specific symptoms), so that this is not a measure of validity but rather a measure of (i) the degree to which subjects can be relied on to recall significant episodes and (ii) the extent to which episodes can be distinguished from everyday vicissitudes by the subject at the DIS interview and by their collaborative witness. If the results are viewed in this manner, they are encouraging in that subjects and witnesses were usually able to agree on what constitutes a significant depressive episode. However, the subjects and witnesses in this group were generally co-operative, intelligent, well-informed adults who seemed to be able to understand

---

Table 9.2    Consistency of reports from 133/150 subjects who were interviewed in person in 1983, and their witnesses

---

		<u>Subject reports an episode of major depression</u>		Kappa	Sensitivity	Specificity
		Yes	No			
Witness reports an episode of depression coinciding in time and impact with subject's	Yes	23	0	0.93	88.5%	100%
	No	3	107			
		<u>Subject reports an episode of minor or intermittent minor depression</u>		Kappa	Sensitivity	Specificity
		Yes	No			
Witness reports an episode of depression coinciding in time and impact with subject's	Yes	41	3	0.86	89.1%	96.6%
	No	5	84			

---



concepts of depression and psycho-social precedents and may be more reliable than subjects in the general population.

There is a methodological caveat in that those few subjects who failed to provide collaborative witnesses may have been those with poor relationships with family and/or few close friends. These interpersonal factors may have also been associated with higher rates of depression. If this were the case, the results may be biased towards inclusion of those subjects who had confidants (either friends or family) whom they were able to call on as witnesses and these witnesses are likely to give more reliable collaborative data.

### iii) Inter-rater reliability between the 1983 and 1988 raters

Prior to the second DIS assessment (at the ten year follow-up), a preliminary study was undertaken to check the inter-rater reliability between the two independent interviewers who used the DIS during the study (i.e. the author, in 1983 and Ms. Curtain, the lay interviewer in 1988). This study involved 10 volunteers and randomly selected study subjects, and was described in Chapter 6 (Table 6.2). Results showed high inter-rater reliability whether or not DIS-derived depressive categories were reported. Twelve of the subjects in the cohort were also given the DIS by one of the two interviewers (as described in Chapter 6) and the results were similarly highly reliable (Table 6.2).

### Discussion of inter-rater reliability

In 1988, inter-rater reliability for generating DIS-derived diagnoses was very high for a mixed group of normal volunteers and clinically depressed subjects. The inter-rater reliability was determined at interviews where both interviewers were present, alternatively conducting the interview. This is an accepted method of determining inter-rater reliability when using the DIS, as was discussed in Chapter 2. There is a problem in that there are some questions which are omitted following certain

negative answers. These influence the fashion in which the non-interviewing rater reports the data and may artificially increase the reliability of the data.

iv) Comparison of rates for diagnostic groupings for periods prior to 1983, from 1983 and 1988 DIS data

Tables 9.3 to 9.5 report rates for the three diagnostic groupings. First, data from the two independent assessments will be reported together for the 161 subjects who completed the DIS on the two occasions, in 1983 and 1988. Rates for 1983 data for 165 subjects (Table 8.1) have been adjusted accordingly.

Table 9.3 reports comparisons of standardised interval rates for DSM-III major depression assessed first in 1983 and then in 1988. Both raters recorded similar rates with no evidence of sex differences on either occasion. But, for lifetime caseness prior to 1978, there was an 11% drop for females and 100% rise for males between the assessments in 1983 and 1988. For 'new case 1978-83' rates there is a 17% fall in rates for females and 53% fall for males (Table 9.3), while lifetime rates to 1983 had dropped 14% for females and 27% for males. Taken overall, there is a trend for both sexes to report slightly fewer episodes of major depression at the 1988 assessment, but there were no sex differences in rates of major depression cross-sectionally as obtained by either rater.

Rates for DSM-III cases (Table 9.4) at both 1983 and 1988 assessments are similar to those for major depression alone due to the small contribution made by reported episodes of dysthymia. There were no sex differences in rates at either DIS assessment. Comparison of data from 1983 to 1988 DIS assessments (Table 9.4) for lifetime rates prior to 1978 showed a 25% drop for females and 100% rise for males; for 'new case rates 1978-83', falls of 3% for females and 62% for males and, for lifetime rates to 1983, falls of 12% for females and 38% for males.

Table 9.3 New cases (rates) and lifetime cases (rates) of DSM-III major depressive disorder assessed by two independent raters at 1983 and 1988

Diagnosis	year DIS given	Female (n=108)	Male (n=53)	Total (n=161)	$\chi^2$
<u>DSM-III major depressive disorder</u>					
<u>1983 DIS assessment - rater K.W.</u>					
Lifetime cases to 1978 (1983)		9 ( 8.3%)	2 ( 3.8%)	11 ( 6.8%)	0.56
New cases 1978-83 (1983)		12 (12.1%)	9 (17.6%)	21 (14.0%)	0.46
Lifetime cases to 1983 (1983)		21 (19.4%)	11 (20.8%)	32 (19.9%)	0.04
<u>1988 DIS assessment - rater R.C.</u>					
Lifetime cases to 1978 (1988)		8 ( 7.4%)	4 ( 7.6%)	12 ( 7.6%)	0.08
New cases 1978-83 (1988)		10 (10.0%)	4 ( 8.2%)	14 ( 9.4%)	0.00
Lifetime cases to 1983 (1988)		18 (16.7%)	8 (15.1%)	26 (16.1%)	0.07
Lifetime case rates = $\frac{\text{number of subjects with episode to that time}}{\text{number of subjects reviewed}} \times 100\%$					
New case rates = $\frac{\text{number of new cases}}{\text{number of subjects reviewed - previous cases}} \times 100\%$					

Table 9.4 New cases (rates) and lifetime cases (rates) of combined DSM-III categories assessed by two independent raters at 1983 and 1988

Diagnosis	Year DIS given	Female (n=108)	Male (n=53)	Total (n=161)	$\chi^2$
<u>DSM-III Cases (MDD and dysthymia)</u>					
<u>1983 DIS assessment - rater K.W.</u>					
Lifetime rates to 1978	(1983)	12 (11.1%)	2 ( 3.8%)	14 ( 8.7%)	1.58
New cases	1978-83 (1983)	14 (14.6%)	11 (21.6%)	25 (17.0%)	0.81
Lifetime rates to 1983	(1983)	26 (24.1%)	13 (24.5%)	39 (24.2%)	0.00
<u>1988 DIS assessment - rater R.C.</u>					
Lifetime cases to 1978	(1988)	9 ( 8.3%)	4 ( 7.6%)	13 ( 8.1%)	0.00
New cases	1978-83 (1988)	14 (14.1%)	4 ( 8.2%)	18 (12.2%)	0.61
Lifetime cases to 1983	(1988)	23 (21.3%)	8 (15.1%)	31 (19.3%)	0.88
<div style="text-align: right;">* p &lt; .05</div> Lifetime case rates = $\frac{\text{number of subjects with episode to that time}}{\text{number of subjects reviewed}} \times 100\%$					
New case rates = $\frac{\text{number of new cases}}{\text{number of subjects reviewed - previous cases}} \times 100\%$					

Table 9.5 New cases (rates) and lifetime cases (rates) of combined RDC categories assessed by two independent raters at 1983 and 1988

Diagnosis	year DIS given	Female (n=108)	Male (n=53)	Total (n=161)	$\chi^2$
RDC Cases (major, intermittent and episodic minor depression)					
1983 DIS assessment - rater K.W.					
Lifetime cases to 1978	(1983)	24 (22.2%)	9 (17.0%)	33 (20.5%)	0.60
New cases 1978-83	(1983)	23 (27.4%)	12 (27.3%)	35 (27.3%)	0.00
Lifetime cases to 1983	(1983)	47 (43.5%)	21 (39.6%)	68 (42.2%)	0.22
1988 DIS assessment - rater R.C.					
Lifetime cases to 1978	(1988)	20 (18.5%)	8 (15.1%)	28 (17.4%)	0.59
New cases 1978-83	(1988)	27 (30.7%)	6 (13.3%)	33 (24.8%)	4.80*
Lifetime cases to 1983	(1988)	47 (43.5%)	14 (26.4%)	61 (37.9%)	4.42*
<div style="text-align: right;">* p &lt; .05</div> Lifetime case rates = $\frac{\text{number of subjects with episode to that time}}{\text{number of subjects reviewed}} \times 100\%$					
New case rates = $\frac{\text{number of new cases}}{\text{number of subjects reviewed - previous cases}} \times 100\%$					

Rates for RDC cases at both assessments are noted in Table 9.5. Comparing rates at 1983 assessment to those at 1988 assessment, for lifetime rates prior to 1978, there was a 17% fall for females and 11% fall for males; for 'new case rates 1978-83', there was a rise of 11% for females and fall of 51% for males; for lifetime rates to 1983, the rates for females were unchanged, while rates for males dropped by 33%. There are statistically significant sex differences in 'new case rates (1978-83)' and lifetime rates to 1983 in the 1988 assessment data that had not been apparent when data for those time periods had been collected in 1983. When new case rates from the two interviews are compared, reported rates for females remained more stable than for males, although this does not imply that the same subjects were classified as cases on each occasion.

Comparisons were made of the numbers (rates in parentheses) for lifetime cases recorded by the two interviewers for females and males respectively. As there were different denominators used for 1983 and 1988 DIS-derived data for calculation of new case rates, these were not compared. For females (Table 9.6), there were no significant differences in rates generated by either interviewer. For males (Table 9.7), there was a trend for rates of lifetime cases to 1983 to be higher when assessed in 1983 than in 1988, most distinct for DSM-III and RDC cases.

A general trend has been noted for both sexes to be less likely to report depressive episodes when reassessed in 1988, for DSM-III major depression and DSM-III case groupings, with males reporting lower rates than females at interview five years later. For RDC cases, there were no overall changes in rates for females but changes in rates of reporting were evident for males. Thus the choice of the diagnostic grouping and sex of the subject had some influence on the magnitude of the change in reported rates over time as reflected in the data reported in Tables 9.6 and 9.7.

Table 9.6      Comparison of numbers (rates) for new cases and lifetime cases for the three depressive groupings, derived from the interviews at 1983 and 1988 by two independent raters, for females

Numbers (rates) for females (n=108)			
Diagnosis	1983	1988	$\chi^2$
<u>DSM-III major depression</u>			
Lifetime cases to 1978	9 ( 8.3%)	8 ( 7.4%)	0.00
Lifetime cases to 1983	21 (19.4%)	18 (16.7%)	0.13
<u>DSM-III cases</u>			
Lifetime cases to 1978	12 (11.1%)	9 ( 8.3%)	0.21
Lifetime cases to 1983	26 (24.1%)	23 (21.1%)	0.11
<u>RDC cases</u>			
Lifetime cases to 1978	24 (27.4%)	20 (18.5%)	0.26
Lifetime cases to 1983	47 (43.5%)	47 (43.5%)	0.00
Lifetime case rates = $\frac{\text{number of subjects with episode to that time}}{\text{number of subjects reviewed}} \times 100\%$			

Table 9.7      Comparison of numbers (rates) for new cases and lifetime cases for the three depressive groupings, derived from the interviews at 1983 and 1988 by two independent raters, for males

Diagnosis	Numbers (rates) for males (n=53)		
	1983	1988	$\chi^2$
<u>DSM-III major depression</u>			
Lifetime cases to 1978	2 ( 3.8%)	4 ( 7.6%)	0.18
Lifetime cases to 1983	21 (19.4%)	18 (16.7%)	0.58
<u>DSM-III cases</u>			
Lifetime cases to 1978	2 ( 3.8%)	4 ( 7.6%)	0.18
Lifetime cases to 1983	13 (24.5%)	8 (15.1%)	1.49
<u>RDC cases</u>			
Lifetime cases to 1978	9 (17.0%)	8 (15.1%)	0.00
Lifetime cases to 1983	21 (39.6%)	14 (26.4%)	2.09
Lifetime case rates = $\frac{\text{number of subjects with episode to that time}}{\text{number of subjects reviewed}} \times 100\%$			



The reported percentage changes may be misleading in that they give no indication as to whether the same subjects were 'cases' or 'non-cases' on each assessment occasion. Further analyses of consistency in pre-1983 lifetime data were then undertaken.

Data gathered by the author in 1983 were compared with data gathered by Ms Curtain in 1988, for the 108 females and 53 males who completed the assessment on both occasions. Kappa values are given both for subjects who were agreed to be 'cases' (Table 9.8), and for individual episodes established at both assessments (Table 9.9).

Kappa values for caseness identified at any time up to 1983 for females ranged from 0.50 for DSM-III cases to 0.62 for RDC cases and, for males, from 0.34 for DSM-III cases to 0.71 for RDC cases (Table 9.8). There was little difference in the ranges of kappa values (0.48 to 0.65 for females and 0.32 to 0.61 for males) when individual episodes were considered (Table 9.9). Categorisation of RDC cases tended to be the most consistent as there is more accommodation for continuing inclusion of the subject as cases in spite of change in the number of symptoms. Overall, the rates for females tended to be more consistent than rates for males, in line with the sex differences in rates reported in Tables 9.6 and 9.7.

#### Discussion of comparison of rates for diagnostic groupings for periods prior to 1983, from 1983 and 1988 DIS data

There is an expectation that data collected with a structured case-finding instrument enquiring after discrete episodes of depression would be more consistent over time than self-report symptom questionnaires, which proved to be the case. Kappa values of 0.53 for major depression compared favourably to those of 0.48 to 0.87 quoted for inter-rater reliability using SADS/RDC criteria for major depression in the NIMH multi-centre collaborative project over periods of weeks (Spitzer and Endicott,

Table 9.8 Consistency of subjects being identified as 'lifetime cases' to 1983, comparing data gathered by two independent raters, five years apart (in 1983 and 1988) for 108 females and 53 males assessed on both occasions

"CASE" - YES OR NO					
Diagnostic Category	YES 1983 YES 1988	YES 1983 NO 1988	NO 1983 YES 1988	NO 1983 NO 1988	Kappa
<u>Females</u>					
Major depression	12	9	6	81	0.53
DSM-III case	15	11	8	74	0.50
RDC case	37	10	10	51	0.62
<u>Males</u>					
Major depression	5	6	3	39	0.43
DSM-III case	5	8	3	37	0.34
RDC case	14	7	0	32	0.71
Major depression = DSM-III major depression (significant)					
DSM-III case = DSM-III major depression or dysthymia					
RDC case = RDC major depression or intermittent minor or minor depression (significant)					

Table 9.9 Consistency of identification of individual lifetime episodes to 1983 assessed by two independent raters, five years apart (in 1983 and 1988) for 108 females and 53 males assessed on both occasions

EPISODE - YES OR NO					
Diagnostic Category	YES 1983 YES 1988	YES 1983 NO 1988	NO 1983 YES 1988	NO 1983 NO 1988	Kappa
<u>Females</u>					
Major depression	17	13	7	81	0.52
DSM-III case	20	17	9	74	0.48
RDC case	48	12	9	51	0.65
<u>Males</u>					
Major depression	7	8	5	39	0.38
DSM-III case	7	10	5	37	0.32
RDC case	17	9	2	32	0.61
Major depression = DSM-III major depression (significant)					
DSM-III case = DSM-III major depression or dysthymia					
RDC case = RDC major depression or intermittent minor or minor depression (significant)					

1987) and 0.20 to 0.75 for major depression over a six-month period (Andreasen et al, 1981). While there is some lack of consistency in reporting pre-1983 episodes, the overall consistency kappas of 0.52 (females) and 0.38 (males) for major depression are clearly superior to those reported by Helzer et al (1985) and Anthony et al (1985), of 0.33 (over three months), and 0.25 (over one month), respectively. This is particularly relevant when, in those studies, the test-retest interval was relatively brief as against the five-year interval in this study.

Wittchen et al (1985) found high agreement in a 7-year follow-up of 158 community subjects between a psychiatrist using ICD-9 and a psychologist generating DIS/DSM-III diagnoses. However, the fact that ICD-9 diagnoses do not demand minimum thresholds for symptom numbers and duration of episodes would tend to increase kappa values as entry to a particular category is not as rigorous and, in the sample of 130 subjects, 42 were omitted because they had been given ICD-9 diagnoses (mainly for personality disorder) that were not covered by DIS categories.

The combined RDC depression categories (RDC cases) grouping in the teachers' study obtained a higher kappa value than the DSM-III major depression or combined DSM-III categories (DSM-III cases) groupings. This is a result of the same phenomenon as noted for Wittchen's comparison of DIS/DSM-III and ICD-9 categories. Wittchen's group (1989) also later reported a study of test-retest reliability over one to two days using a different rater (alternately a psychiatrist or psychologist) on each occasion. Here they reported a fall-off in symptoms and number of episodes in most DSM-III categories for recent episodes but much greater consistency for lifetime diagnosis rates.

In another recent study (McLeod et al, 1990), 353 (73%) of 484 subjects who had been interviewed by lay interviewers using the DIS in a community sample in Detroit, U.S.A., were re-interviewed by a team of clinicians (psychiatric social

workers). The median lag time to second interview was 11 weeks. The DIS interviews generated RDC diagnoses for individual and combined depressive (RDC case) categories for the six-month period prior to the initial interview. The authors reported kappas of 0.39 for all RDC cases, 0.28 for major depression, 0.10 for intermittent minor and 0.08 for minor depression. They noted that longer, severe episodes were more consistently recalled and that inconsistency in reporting was most often associated with the timing of the episode (i.e. whether or not the episode had occurred within the six month span). They considered such issues as differences in empathy between clinical and lay interviewers, slight differences in wording of the questions at each interview and cue effects from the original interview but concluded with the opinion that the major cause of differences was inconsistent recall by the subjects. These authors have used psychiatric social workers as the clinician raters and the study raises a question as to the comparability of categorical diagnoses derived by psychiatrists and psychologists.

None of the studies mentioned have entertained a discussion of sex differences in recall. However, the sex difference issue had previously been raised in a study of a cohort of 591 young Swiss adults (Angst & Dobler-Mikola, 1984c). Here prevalence rates for DSM-III and RDC depressive episodes generated by use of their own semi-structured interview schedule (SPIKE) varied with the time period used, so that there was an equal sex ratio for three-month prevalence and female preponderance for twelve-month prevalence, which the authors attributed to males' more readily forgetting episodes that were further removed in time.

The five studies summarised raise a number of methodological issues. The first two studies emphasized differences between lay and professional interviewers, the latter three highlight respondent recall error. The consistency of recall may be determined by the motivation of both interviewer and subject, time lag between interviews and period of time being studied. Ironically, the use of a short time period may increase inconsistency of recall of timing of episodes due to the actual episodes being recalled

but the timing being out by a factor of weeks. Weissman and colleagues (1988) have questioned whether the DIS, as a stand-alone instrument is sufficient to prompt recall for accurate estimates of lifetime prevalence, and other authors (Wittchen 1989, McLeod 1990) have recently suggested use of such cues as birthdays and holidays to increase accuracy of recall.

v) Comparison of reported numbers of symptoms and duration of episodes from 1983 and 1988 assessments

Comparison was made of the mean number of symptoms for the period prior to 1983 reported to the psychiatrist rater at 1983 with the mean number for the same period reported to the lay interviewer in 1988. Comparisons involved the group assessed both at 1983 and 1988 (i.e. 108 females and 53 males). When it was clear that the same pre-1983 episode or episodes were reported at both assessments (in 1983 and 1988), the number of symptoms reported on each occasion for each episode were compared. Here two diagnostic groupings, major depression (representing the highest diagnostic threshold in terms of symptom numbers) and RDC cases (representing the lowest diagnostic threshold in terms of symptom numbers) were assessed. Dysthymia was only considered in the combined RDC categories as there were generally low numbers of episodes which were inconsistently recalled (particularly by males), which had the potential to lead to great variation in results when considering sex differences in duration of episodes.

There were no significant sex differences in the mean number of symptoms reported for both DSM-III major depression and RDC case groupings, but females reported fewer symptoms per episode five years later (in 1988), while males did not (Table 9.10).

The episodes reported by males on both occasions were of longer mean duration for episodes of DSM-III major depression when 1983 and 1988 assessments were

Table 9.10      Number of symptoms for pre-1983 episodes recorded by both interviewers, in 1983 and 1988

Number of episodes		Number of symptoms Mean (SD)	Number of symptoms Mean (SD)	t-test
Year		1983	1988	
<u>Type of episode</u>				
<u>DSM-III major depression</u>				
Female	16 episodes	5.94 (1.18)	4.88 (1.02)	3.78**
Male	7 episodes	5.71 (1.25)	5.00 (1.24)	1.37
t-test		0.40	-0.25	
<u>RDC cases (combined RDC categories)</u>				
Female	52 episodes	3.90 (1.67)	3.44 (1.29)	3.00**
Male	17 episodes	4.29 (1.61)	3.94 (1.35)	1.06
t-test		-0.86	-1.34	
				** p < .005

Table 9.11 Duration of episodes for episodes (considered present by two independent raters) and reported at 1983 and 1988 for 108 females and 53 males assessed on both occasions

Number of episodes		Duration in weeks Mean (SD)	Duration in weeks Mean (SD)	t-test
Year		1983	1988	
<u>Type of episode</u>				
<u>Major depression</u>				
Female	16 episodes	16.44 (20.41)	12.31 ( 9.46)	0.95
Male	7 episodes	41.14 (16.49)	37.14 (22.24)	0.51
t-test		-3.07*	-2.84*	
<u>RDC cases (combined RDC categories)</u>				
Female	52 episodes	27.77 (37.75)	22.73 (35.19)	1.17
Male	17 episodes	33.94 (47.34)	26.94 (20.91)	0.66
t-test		-0.49	-0.60	
				* p < .05



compared (Table 9.11). There were no significant sex differences in duration for combined RDC categories.

Discussion of comparison of reported numbers of symptoms and duration of episodes from 1983 and 1988 assessments

Both males and females reported a mean of 5-6 symptoms from a maximum of 8 symptoms per episode of major depression at the 1983 and 1988 assessments (Table 9.10). It is worth noting that when episodes of episodic and intermittent minor depression are added to major depression (i.e. RDC cases), the mean number of symptoms per episode was consistently around four so that these episodes are not trivial in terms of reported distress. The mean duration of episodes of DSM-III major depression ranged from 12 to 41 weeks (Table 9.11). It is more difficult to comment on the duration data, as the mean duration of episodes is confounded by the discrepancy in minimum duration between episodic (2 weeks) and intermittent minor (2 years) depression.

There are no sex differences in reported symptoms in the episodes of major depression reported at both 1983 and 1988. There was a statistically significant decrease in reported symptoms per episode over time (from 1983 to 1988) for females only. The same patterns are reported for the combined RDC categories, where episodes of major depression still make a significant impact on mean scores for both symptom numbers and duration.

The possible explanations for the overall decrease in number of symptoms over time are (i) a practice effect, (ii) a function of change in attribution of depressive episodes by the subjects, (iii) diminished recall by subjects over time, or (iv) differences in interviewing style.

Of the possible explanations, it is unlikely that a practice effect is operating due to the long interval between assessments. Furthermore, differences in interviewing style were kept to a minimum by the use of the DIS. The two other possible explanations both involve differences in subjective recall over time. While reported duration of episodes decreased for both sexes, males reported longer duration per episode for those episodes that were recalled at both assessments. When the finding that females reported less symptoms per episode over time than males is coupled with the finding of longer mean duration for episodes recalled on both occasions (Table 9.11) in males, a possible explanation is that males tended to only recall more severe episodes (in terms of duration) and which were likely to be recalled more reliably in terms of severity than those recalled by females (from a broader range of severity).

The sex differences in rates at the 1988 assessment are consistent with the faster rates of forgetting in young males reported by Angst's group (1984c) but suggest that these sex differences may be prominent for episodes of lesser severity.

vi) Consideration of various methods of estimating lifetime case rates using data from 1983 and 1988 DIS assessments

The two independent vantage points in time can be used to derive lifetime prevalence rates to 1988 using three different methods, reported in Table 9.12.

For the '1988 assessment' figures, the data refer only to data collected by the lay interviewer in 1988 and have already been reported (in Table 8.2). A second or 'combined estimate' method respects caseness decisions made at both the 1983 and 1988 DIS interviews. A third or '1983/88 add-on estimate' method respects the decisions (i.e. 'case'/'not case') made in 1983 by the author. The 'new case' diagnoses from 1983-88 time period are then added to the 1983 rates as though the 1988 lay rater, at the 1988 DIS interview, only enquired about depressive experience for the 1983-88 interval. Here a 'new case' in the 1983-88 time period is a subject who reported a first depressive

Table 9.12 Lifetime cases (rates) to 1988 determined using three different methods of estimation derived from 1983 and 1988 DIS data

Diagnosis	year DIS given	Female (n=108)	Male (n=53)	Total (n=161)	$\chi^2$
<u>Major depression only</u> (DIS significant DSM-III major depressive disorder)					
1988 estimate	1988	28 (25.9%)	9 (17.0%)	37 (23.0%)	1.61
'Combined estimate'	1983+1988	35 (32.4%)	14 (26.4%)	49 (30.4%)	0.35
1983/88 estimate	1983/88	29 (26.9%)	12 (22.6%)	41 (25.5%)	0.33
<u>DSM III-case</u> (major depressive disorder and dysthymia)					
1988 estimate	1988	33 (30.6%)	10 (18.9%)	43 (26.7%)	2.48
'Combined estimate'	1983+1988	41 (38.0%)	17 (32.1%)	58 (36.0%)	0.53
1983/88 estimate	1983/88	35 (32.4%)	15 (28.3%)	50 (31.1%)	0.28
<u>Combined RDC categories</u> (major, intermittent and episodic minor depression)					
1988 estimate	1988	61 (56.5%)	19 (35.8%)	80 (49.7%)	6.05*
'Combined estimate'	1983+1988	65 (60.2%)	22 (41.5%)	87 (54.0%)	4.99*
1983/88 estimate	1983/88	58 (53.7%)	22 (41.5%)	80 (49.7%)	2.15
					* p < .05
1988 estimate	= cases taken from 1988 DIS assessment only				
'Combined estimate'	= cases at 1983 and/or 1988 assessments				
1983/88 estimate	= cases to 1983 (1983 assessment), then new cases for 1983-88 (1988 assessment)				

episode in the time period 1983-88 (at 1988 DIS interview) and had not previously been a 'case' using the data derived from the 1983 DIS interview.

For the '1988 estimate' method, the RDC lifetime case rate (Table 9.6) shows a significant female preponderance (57% vs 36%) while a similar non-significant trend (31% vs 19%) for DSM-III cases. As the suggested female preponderance is not striking for major depression alone (26% vs 17%) it becomes more apparent as the threshold for caseness is lowered by inclusion of episodic and intermittent minor depressive disorders.

Using the second or 'combined estimate' method, lifetime rates for RDC cases to 1988 are 60% for females and 42% for males, with a smaller but still statistically significant sex difference. Lifetime rates for DSM-III cases are now 38% for females and 32% for males, and for major depression alone, 32% for females and 26% for males, respectively.

Using the third '1983/88 estimate' method, the lifetime rates for RDC cases are now 54% for females and 42% for males with a trend towards a female preponderance, which is no longer statistically significant. Lifetime rates for DSM-III cases are now 32% for females and 28% for males, for major depression alone, 27% for female and 23% for males. Using the third method, the sex ratios for all categories have diminished and there is a non-significant trend towards female preponderance in the combined RDC categories.

#### Discussion of estimation of lifetime prevalence rates

The use of the '1988 estimate' method leads to the greatest female preponderance in rates for all diagnostic groupings, which is statistically significant for the RDC case grouping. Data reported in Table 9.12 show that the RDC grouping is the

most likely system to show sex differences (with statistically significant differences for two of the three methods and a trend being apparent for the '1983/88 estimate' method).

The discussion in previous sections has considered the possibility of sex differences in rates of recall up to 1983. The lifetime rates to 1988 will also be determined by the 'new case rates' in the 1983-88 period (Table 8.2) which will now also be considered.

There are possible biases in each of the methods. The '1988 estimate' has a potential sex bias as those males who reported earlier episodes (at 1983 interview) which were not recalled at 1988 interview will not be noted as cases unless they reported subsequent "new" episodes in the 1983-88 period, while females were more likely to recall episodes previously reported. For the 'combined estimate' method, the males who "forgot" are retained as cases but there is still a potential bias towards female preponderance as some females tended to "remember" previous episodes while no males did. This method is the most inclusive, allowing any episode reported at either 1983 or 1988 assessment. The '1983/88 estimate' assumes that the 1983 interview, supported by witness reports, generated more reliable data for the period to 1983 than the 1988 assessment. This method retains as cases those males who "forgot" episodes reported earlier (at 1983) when asked at the 1988 interview but does not include those females who subsequently recalled episodes (at 1988) that had not been previously reported (at 1983), and has less potential sex bias due to different rates of recall for males and females.

The rates for lifetime cases using all three methods (see Table 9.12) report similar trends towards female preponderance, but there are no significant sex differences for DSM-III categories of major depression with or without dysthymia. For combined RDC categories, the female preponderance is increased due to sex differences in reporting of rates of minor depressive episodes over time. The magnitude of the sex

ratio is affected by decisions as to which estimate is accepted as providing the most reliable data.

The sex differences are minimised using the 1983/88 lifetime rates, as males who later "forgot" are still included, but females who later "remembered" are not.

Two explanations of the differing rates of female preponderance in RDC lifetime case rate (being greatest for the '1988 estimate') require exploration. Firstly, this may be a valid finding reflecting a gradually increasing (but generally statistically non-significant) female preponderance, most marked in the new case rate from 1983 to 1988. There are trends (Tables 9.3 to 9.5) to support that interpretation. Secondly, that there had been a real female preponderance in 1983 but that had not been established at the prior (1983) interview, either because of difference in raters, or statistical explanations, such as low base rates or Type II error.

The second explanation is doubted as high inter-rater reliability and moderate agreement on caseness and individual episodes has been shown (particularly for the pre-1978 period), which would argue against decay of recall related simply to passage of time. The possibility of a Type II error is dealt with in Chapter 11, but this issue is just as pertinent on both 1983 and 1988 data.

The first explanation will be explored further in the context of factors influencing reliability of the data and sex differences in recall.

vii) Further examination of collaborative witness reports at 1983 interview to examine disparities in 1983 and 1988 DIS data

The collaborative witness interview in 1983 was described earlier in the chapter. The witness accounts were examined with particular reference to subjects where there

was disparity between reported episodes prior to 1983 from the 1983 and 1988 DIS assessments.

For DSM-III cases, there were 11 females who had been categorised as cases in 1983 but not in 1988. Of these, 10 had provided witnesses and all the witnesses had reported episodes of depression which were similar to those described by the subject. In one situation, the witness described an extra episode not reported by the subject. These were similarly 8 males, of whom 7 provided witnesses. Six of the witnesses reported episodes that were similar to those reported by the subject and one reported no episodes. In the latter case, the subject had reported seeking help and taking medication for the episodes (at 1983 interview).

Where episodes were reported at 1988 assessment that had not been noted at 1983, for DSM-III cases, there were 8 females, 5 of whom had provided witnesses. In three subjects, a diagnosis of minor depression had been made and witnesses had also noted episodes of depression (so that the discrepancy was due to reporting of more symptoms at 1988 than 1983); in two situations, witnesses at 1983 also had not reported episodes (thereby agreeing with the subject's report in 1983). The three males had all provided witnesses who had reported episodes at 1983 which coincided with the reports of the subjects.

For RDC cases, there were ten females who reported episodes at 1983 but not at 1988, for the pre-1983 time periods. Of these, 8 provided witnesses, of whom 6 gave comparable episodes to the subject (with one witness again giving extra episodes) and two witnesses reported the same precipitating incidents as the subject but did not regard the ensuing distress as constituting episodes of depression. Since 1983, three of the 10 females had gone on to report further episodes of major depression; three reported further episodes of minor depression and three reported the onset of panic disorder. All 7 males provided collaborative witnesses who all gave similar accounts of depressive

episodes as the subjects. Since 1983, one male reported a further episode of major depression, 3 reported episodes of minor depression and 2 reported onset of generalised anxiety disorder.

There were 10 females who had not reported episodes of depression for the period prior to 1983 at 1983 assessment but then did so in 1988. Of these 10, 8 provided witnesses and in all cases, the witness at 1983 had not volunteered any episodes of depression. One of these subjects had reported an episode of distress following a marital breakup and had sought advice from her general practitioner. Another subject had reported a time of unhappiness which the witness also reported as an episode of distress and the subject had reported seeking advice medication from their general practitioner. This subject later reported an episode of major depression in 1983-88 period. None of the other seven who provided witnesses had reported any further episodes of depression in the 1983-88 period.

#### Discussion of collaborative witness data in light of disparities between 1983 and 1988 data

When examining data for episodes of major depression, the major source of discrepancy was a change in the number of symptoms reported, so that episodes were categorised as major depression on one occasion and minor depression on another. In the subjects, earlier recall of episodes of major depression may have been impaired by the later onset of panic disorder.

Episodes of dysthymia were the least reliably recalled. Here, some episodes were omitted at later interview (particularly for males) and other episodes were recalled with a shorter duration, so that these episodes were then categorised as minor depression but subjects who had previously been 'DSM-III cases' were then RDC cases only. As the most of the episodes of dysthymia were initially reported as lasting between two and three years, the shorter reported duration became an important source



of unreliability. Earlier reference has been made to the heterogenous nature and lack of clear conceptualisation of this category.

For RDC cases, any subjects who had reported episodes as major depression or dysthymia on one occasion and minor depression at the other occasion were all still categorised as cases on both occasion. Here an important source of error was the reporting of a reaction to a life event as distress on one occasion but then reporting more symptoms and 'significant impact on life' at the other interview, so that a diagnosis of at least minor depression was then made. This was an issue that only involved females and, on examination of the collaborative witness data, the witnesses tended to agree with the subjects' earlier assessments. On the basis of this finding, the lifetime prevalence rates based on the 1983/88 assessment method (i.e. diagnoses made prior to 1983 (from 1983 assessment) and from 1983 to 1988 (from 1988 assessment) will be used for determination of risk factors to lifetime caseness in Chapter 10.

#### Data concerning recourse to help-seeking

All subjects were asked at each of the three assessment points whether they had sought professional help for depression or reported receiving prescription of psychotropic medication for "nerves". The questions concerning help-seeking and medication used were asked in the semi-structured interview prior to the DIS. This does not simply involve a reporting of data derived from DIS probe questions, which would confound the caseness rates (as entry to depressive categories may be predicated on such responses). These data are shown in Table 9.13 in terms of sex differences. Reports of help-seeking and medication for "nerves" at 1978, 1983 and 1988 were then compared with the numbers of subjects categorised as DSM-III or RDC lifetime cases (Table 9.14).

These tabulated data allow two broad conclusions. Firstly, while there are no sex differences in help-seeking patterns or in receiving psychotropic medication, there

Table 9.13 Professional treatment details reported at any time prior to at the 1978, 1983 and 1988 assessments

Treatment variable	Assessed	Number (rate)		$\chi^2$
		Female	Male	
Received medication "for nerves"	to 1978	13 (12.4%)	3 (5.7%)	1.75
	to 1983	27 (25.0%)	10 (18.9%)	0.78
	to 1988	27 (25.0%)	12 (23.6%)	0.07
Consulted professional (e.g. local doctor psychiatrist, psychologist)	to 1978	15 (14.6%)	4 (7.7%)	1.52
	to 1983	28 (25.9%)	13 (24.5%)	0.04
	to 1988	42 (38.9%)	14 (26.4%)	2.44

Table 9.14		Level of agreement between caseness estimates and subjects' chance of consulting a professional, over three lifetime estimates (taken at 1978, 1983 and 1988)					
		<u>Consulted Professional</u>			Kappa	Sensitivity	Specificity
			Yes	No			
RDC 'case'	1978	Yes	10	25	0.25	82%	53%
		No	9	112			
	1983	Yes	33	35	0.42	48%	91%
		No	8	85			
	1983R <sup>+</sup>	Yes	27	34	0.32	44%	86%
		No	14	86			
	1988	Yes	46	34	0.45	57%	88%
		No	10	71			
DSM-III 'case'	1978	Yes	5	9	0.22	93%	26%
		No	14	128			
	1983	Yes	21	18	0.37	54%	84%
		No	20	102			
	1983R <sup>+</sup>	Yes	17	14	0.32	55%	81%
		No	24	106			
	1988	Yes	29	14	0.41	67%	77%
		No	27	91			
<sup>+</sup> The second 1983 estimate was a retrospective one, derived by rater R.C. in 1988.							

is a trend for the consultation rate for females to be higher by 1988 that may indicate that the female preponderance is a real finding. Assuming that professional consultation for depression and status as a lifetime depressive 'case' correspond, examination of the level of agreement should be a further check on case rate data. Table 9.14 shows low levels of agreement between report of professional consultation prior to 1978 and RDC/DSM-III caseness (kappas of 0.25 and 0.22). The kappas comparing 'immediacy' data (i.e. pre-1983 assessments undertaken in 1983; pre-1988 assessments undertaken in 1988) show moderate levels of agreement. The sensitivity and specificity estimates are encouraging, in that subjects allocated as lifetime 'cases' at each assessment were reasonably likely to have sought professional assistance, while 'non-cases' were less likely to have received professional assistance. The 1988 retrospectively derived pre-1983 data were less impressive than the pre-1983 data derived at the 1983 assessment. On further examination of the data, those categorised as 'non-cases' who had sought professional help had approached their general practitioner concerning distress over break-up with relationships rather than psychiatric consultation. On the other hand, the relationship between early onset episodes of major depression reported to both interviewers and help-seeking from psychiatrists has already been noted.

#### General discussion of reliability of DIS-derived data

What factors might have determined the inconsistency in reporting pre-1983 lifetime data, assessed initially in 1983 and subsequently in 1988? Firstly, did the second (non-psychiatrist) rater underestimate generally in comparison to the author (the psychiatrist rater)? This is unlikely, as the lay rater returned similar total case rates in 1988 of 42.2% (RDC) and 19.3% (DSM-III), as against the psychiatrist rater's total case rates in 1983 in 1988 of 42.2% (RDC) and 24.2% (DSM-III). The rates were almost identical for females but lower for males.

A sub-analysis restricted to those who reported the same pre-1983 lifetime episode in 1983 and in 1988, showed clear evidence of fewer symptoms of major depression being remembered after a longer interval and, in any such community sample, this would tend to move subjects from "case" to "non-case" status, contributing to the change to non-caseness of 20% for the DSM-III and of 10% for the RDC system (calculated from Table 9.4 and 9.5 data). Any general trends to forget depressive episodes and/or specific episode features was balanced by females (but not males) to report pre-1983 episodes in 1988 which they did not report in 1983, suggesting that females were either more likely to remember earlier forgotten episodes or to report more episode features so that they subsequently achieved 'case' status.

There is no reason to believe that the professional status or training deficits of the two interviewers contributed to the inconsistency. Random interviews of cohort subjects by the two raters established consensus in diagnostic allocation, while at the same time, several subjects actually changed their diagnostic category for previously reported episodes. Thus, this analysis suggested that, given the same information, the two raters came to the same categorical diagnostic decision and that variation in diagnosis-generating data emerged from information presented by the subjects.

Thus the reliability over time is moderate at least and better than reliability between lay and clinician interviewers over shorter durations in some cases. When discussing differences between lay and clinical interviewers, it should be noted that there is considerable variations in the credentials of both groups (Eaton et al, 1984). Lay interviewers generally (but not necessarily) have university degrees and, for the American studies, are often recent graduates or students on university holidays (Eaton et al, 1984). Clinician interviewers in the American studies are also often psychiatrists who have recently finished their training (e.g. Anthony et al, 1985), or are still in training (Helzer et al, 1985), while in the test-retest reliability studies for the PSE, the clinicians tend to be far more experienced (Wing et al, 1977).

At the time of the 1983 DIS interview, the author had twelve years' psychiatric experience and Ms. Curtain, the lay interviewer, was a recent graduate in Psychology (as well as a trained nurse) and was working full-time on the project for twelve months. Both interviewers in this study had more clinical experience and longer-term commitment to the project than the interviewers on the ECA reliability trials while the homogeneity and co-operation of the subjects have already been noted. These factors, coupled with the discrete nature of episodes in these subjects (who were generally not engaged in chaotic life-styles nor prey to chronic stressors such as long-term poverty) may explain the moderate, but satisfactory, reliability of the data over a long (five year) follow-up period.

The corroborative witness interviews are supportive of the pre-1983 lifetime estimates in 1983 and of their use in establishing reliability of data where there were differing reports of episodes between 1983 and 1988 interviewers. In 1988, there was no collaborative witness interview but professional consultation was much more likely to be reported for RDC 'cases' [ $\chi^2 = 36.27$ , df 1,  $p < 0.001$ ] and DSM-III 'cases' [ $\chi^2 = 27.39$ , df 1,  $p < 0.001$ ]. Thus it seems likely that the estimates of "caseness" (as defined by DSM-III and RDC) were valid, and that a real trend for a female preponderance in depression rates had then emerged, where statistical significance was related to definition of caseness. The inconsistency in pre-1983 estimates suggests, however, that such a conclusion can only be part of the answer at best, and that the emerging sex difference may be determined entirely or in part by an artifact in reporting.

This raises a substantial concern. Imagine that the cohort had been surveyed for the first time in 1988 and depressive episodes had been assessed over the preceding five years and lifetime. The data, and estimates of professional treatment for depression, would have indicated a clear female preponderance, as established in numerous community surveys (see Weissman and Klerman, 1977; Nolen-Hoeksema, 1987). The

cohort study design suggests that, if the female preponderance is a valid finding, then it is less marked in a more socially homogeneous group but may still develop when subjects grow older and/or demonstrate social role divergence, again arguing against exclusive biological determinants. Although the 1988 rates were satisfactory, there is an effect of males being less likely to recall previous episodes at later interview. The findings based on data generated with the DIS on two quite temporally distant occasions indicate that there are likely to be two trends leading to female preponderance. One of these is an artifact effect leading to a perceived decrease over time in rates for males and the other is a possible increase in rates for females more evident in the past few years. The findings are compatible with an artifact effect, with the sexes differing in their reporting of depressive episodes, although the difference is likely to reflect both a real phenomenon and artifactual biases. This issue will be discussed further in Chapter 11.

### Summary

Patterns for responding to self-report questionnaires over a ten-year period showed moderate to high levels of reliability, with state scores predictably demonstrating the lowest long-term reliability and PBI scores showing the highest.

The DIS-derived depression categories showed moderate levels of reliability that were similar to those demonstrated for test-retest reliability trials over much shorter periods. These results were unaffected by whether interviews were carried out in person or by a mailed version with telephone clarification.

The kappa value increased when minor and major depression were considered together as some subjects reported the same episodes with changes in the number of symptoms sufficient to change the category of episode from minor to major depression or vice versa.

Rates for dysthymia proved to be the most inconsistent, particularly for males, although the base rate for this disorder was low.

There were sex difference in rates of recall, which led to statistically significant sex differences in rates when data were gathered after a five-year lapse. The differences become apparent as the threshold criteria for caseness were lowered (i.e. for RDC cases, not for major depression alone or DSM-III cases).

There is support for an artifactual sex bias in forgetting.

Alternative methods of estimating lifetime prevalence rates were discussed, using either (i) 1988 data only, or (ii) all reported cases from 1983 and 1988 data, or (iii) 1983 estimates to 1983 and then adding 1988 estimates for 1983-1988 period. After comparing reports from collaborative witness data at 1983 to accounts of depressive episodes at 1983 and 1988, the 1983 data were found to be more reliable for the pre-1983 period so that the third method of estimating lifetime prevalence was taken as the most reliable.

There was low to moderate agreement between help-seeking patterns and depressive caseness and some support for a trend towards female preponderance over the last five years of the study (1983-1988).

These results are most encouraging but should not be generalised to other population groups too readily as this group are likely to be far more reliable in their patterns of reporting due to the special characteristics of the group. Nevertheless, the findings should suggest the comparative strengths and weaknesses of several case-finding approaches and the utility of a number of 'validating' strategies.



## CHAPTER TEN

### RISK FACTORS PREDICTING CASENESS

#### CONTENTS

Choice of predictor variables  
 The position taken concerning relevance of negative life events  
 Choice of dependent variables  
 Method  
 Discriminant function analysis predicting caseness at 1983  
 and 1988  
 Importance of the risk factors in predicting caseness

Discussion  
 Introduction  
 Neuroticism as a risk factor  
 Dependency as a risk factor  
 Low self-esteem as a risk factor  
 Perception of poor parenting as a risk factor  
 Feminine sex-role characteristics as a risk factor  
 High trait depression as a risk factor  
 Absence of a 'core tie' as a risk factor  
 Presence of three or more children at home as a risk factor  
 Female gender as a risk factor

Conclusions

#### Introduction

The longitudinal nature of the study allows for investigation of risk factors measured prior to onset of depression. These will be pursued using several outcome indicators of caseness. The three diagnostic groupings used to define caseness are (i) DSM-III cases (combined DSM-III categories for affective disorders), (ii) RDC cases (combined RDC categories for affective disorders) and (iii) 'defined' cases. Data concerning rates and reliability for the first two groupings have been reported in Chapters 8 and 9. The 'defined case' grouping was described in Chapter 8 and was introduced to make use of data gathered at more than one assessment (thus increasing reliability) and introduce further operationalisation of impairment criteria that were appropriate for this cohort. The different thresholds for caseness implicit in three diagnostic groupings allow examination of the proposition that the effect of the putative risk factors (especially gender) is influenced by the definition of caseness used.

The aim of the study was to examine the relevance of a number of putative risk factors in discriminating between cases and non-cases using the three caseness definitions.

### Choice of predictor variables

A number of risk factors have been identified in the literature and were included in hypotheses 3 and 4 described in Chapter 6. The choice of predictor variable was determined by the original hypotheses, with further consideration of the performance of putative risk factors in the univariate analyses detailed in Chapter 7. Preference was shown for predictor variables designed to measure trait rather than state effects in order to avoid the confounding effect of depressive state on predictor variables.

#### (i) Self-report measures

**Low self-esteem** (Brown et al, 1986; Robson, 1988) and **high neuroticism** scores (Weissman et al, 1978; Boyce and Parker, 1985) have both been considered as risk factors to depression. There have been concerns about the possible contaminating effects of state depression on self-esteem (Ingham et al, 1986) and on neuroticism measures (Katz & McGuffin, 1987). However mood state has been found to have little effect on EPI neuroticism scores when subjects rate as they "perceive themselves generally" (Boyce and Parker, 1985). **High trait depression** as measured by the Costello and Comrey measure (1967) was included as trait depression scores were shown to be moderately consistent over the ten-year time period (Table 9.1), unlike scores for the state depression measure (Wilson, 1979).

**High dependency** is held to be a key personality risk factor to depression (Hirschfeld et al, 1977). The Depressive Experiences Dependency sub-scale (Blatt et al, 1975) examines the need to rely on, and interact with others when depressed and was included as a dependency measure.

**High scores on the femininity scale** of the Bem Sex Role Inventory (Bem, 1974) quantifies aspects of the feminine sex role which tends to overlap with the 'nurturant' role considered to be a risk factor for depression (Gove, 1984) and allows differentiation between biological sex and femininity as risk factors for female preponderance. It is worth noting that both sexes are rated independently on each scale so that males can score highly on the femininity scale and females can score highly on the masculinity scale.

(ii) Univariate analysis of vulnerability factors from the Brown and Harris model, with consideration of other related psychosocial variables

Four vulnerability factors to depression for women described by the Bedford College team (Brown & Harris, 1977) were (i) presence of a confiding relationship with a husband, boyfriend or lover, (ii) death of mother before age of 11, (iii) presence of three or more children at home under the age of 14, and (iv) lack of employment outside the home. These vulnerability factors were considered to require the presence of adversity prior to onset of depression. However, such adversity is partly controlled for in this cohort in that subjects tended to be experiencing a similar range of life experience (as shown in Chapter 7). At 1983, only one of the four vulnerability factors was of relevance, while at 1988, three factors had relevance to this cohort. The factors will be considered, where relevant, for both sexes.

At the 1983 assessment, the only factor of relevance was **the absence of an intimate partner or, adverse interpersonal characteristics of a partner**, where present. In considering the relevance of intimate relationships, the group was subdivided, allowing comparison of those who had an intimate partner with whom they had a confiding relationship [i.e. 'core tie'] and those who had no such partner.

When subjects reported the presence of an intimate partner, they were asked to complete the Intimate Bond Measure or IBM (Wilhelm & Parker, 1988) to assess the

quality of the relationship. Inspection of the IBM scores, showed a mean score for the care subscale of 30.5 (SD 5.4) at 1983 and 29.5 (SD 6.5) at 1988, approaching the maximum potential score of 36; and on both occasions about 40% of the group had scores of 32 or above, evidencing considerable skewing. By contrast, for the control sub-scale, the mean scores were 6.7 (SD 5.0) at 1983 and 6.9 (SD 6.4) at 1988. Correction for the skewing of the care subscale was made by treating both 'care' and 'control' variables as high/low categorical variables, with putative risk factors being defined as follows: '**low care**' represented a care score of one or more standard deviations below the mean score for IBM care, and '**high control**' represented a score of one or more standard deviations above the mean score for IBM control.

Brown and Harris (1978) had reported the **presence of three or more children under age of 14 years** as a risk factor in working class women. At the 1988 assessment, there were a number of subjects with three or more children under the age of 14 and, increasingly, the females had left full-time work, so that part-time work or lack of paid employment had become reasonably common (as described in Chapter 7). By 1988, 18/108 females (16.7%) and 9/53 males (17.0%) had three or more such children, so that the presence of 3 or more children was used as a categorical predictor variable for both sexes at 1988.

The numbers of cases and non-cases with any number of children were also compared (for the 161 subjects fully reviewed at 1988, using the three caseness definitions). Using the three caseness definitions, 30/50 (or 60%) of DSM-III cases had children and 70/111 (63.1%) of non-cases ( $\chi^2=0.14$ , ns); 49/80 (61.3%) of RDC cases had children and 51/81 (63.0%) of non-cases ( $\chi^2=0.05$ , ns); 26/51 (or 51%) of 'defined' cases and 74/110 (67.3%) of non-cases ( $\chi^2=3.93$ ,  $p < .05$ ), the last analysis being statistically significant. When comparisons were made for those who had three children or more under the age of 14, for both sexes, 7/50 (or 14.0%) of DSM-III cases had 3 or more children compared with 20/111 (18.0%) non-cases ( $\chi^2=0.40$ , ns); 13/80 (16.3%)

of RDC cases and 14/81 (17.3%) of non-cases ( $\chi^2=0.03$ , ns) and 8/51 (15.7%) of 'defined' cases and 19/110 (17.3%) of non-cases ( $\chi^2=0.06$ , ns). When females alone were compared, 4/35 (or 11.4%) of DSM-III cases had three or more children and 14/73 (19.2%) of non-cases ( $\chi^2=1.36$ , ns); 8/58 (13.8%) of RDC cases and 10/50 (20%) of non-cases ( $\chi^2=0.37$ , ns); 5/36 (13.9%) of 'defined' cases and 13/72 18.1%) of non-cases ( $\chi^2=0.51$ , ns). There were no statistically significant differences between cases and non-cases in terms of the presence of three or more children for both sexes or when females were examined alone.

Brown and Harris' (1978) third risk factor concerning early death of mother was of no relevance to this group and was not considered, as only two subjects had lost a parent by death before the age of 11 (in both cases, a father) and a further three subjects had been separated from a father in childhood (by parental divorce).

**Dysfunctional parenting** was not considered by Brown and Harris but was examined here as measured by Parental Bonding Instrument or PBI (Parker et al, 1979) scores. Both maternal and paternal subscales for care and protection from the PBI were entered separately.

Those subjects who had lost one parent by death or parental divorce had only completed PBI scales against their mothers. Two ways of addressing the issue of missing PBI scores for these subjects are by (i) substituting group means for the missing values in the analyses or by (ii) using a multiple regression analysis to predict paternal PBI scores from the maternal PBI scores for those subjects with no paternal scores. After statistical advice from Mr. D. Hadzi-Pavlovic, the second procedure was used to predict paternal PBI scores for the five subjects (3 females and 2 males), who all had maternal PBI scores.

Brown and Harris' fourth risk factor, **being at home without paid employment**, had only limited application to the group studied. Home duties as exclusive employment was only relevant for females and was not common (and therefore not considered) at 1983, while for 1988 data, a separate analysis (with home duties added as an extra variable) was undertaken for females (see Table 10.9).

### The position taken concerning relevance of negative life events

The method of measurement of life event scores was described in Chapter 6 and results for positive and negative life events reported on Table 7.11, where no sex differences were found. Life event score data were available for the 165 subjects who completed the 1983 assessment (for the twelve months prior to 1983). Here scores for negative life events are compared for cases and non-cases using the three caseness definitions. There were statistically significant differences in the means for all three diagnostic groupings. For the DSM-III grouping, means for negative life event scores were 124 (SD 136) for cases and 71 (SD 91) for non-cases ( $t=2.28$ ,  $p < .05$ ); for the RDC grouping, means were 107 (SD 118) for cases and 66 (SD 92) for non-cases ( $t=2.39$ ,  $p < .05$ ); for 'defined' cases, the means were 136 (SD 141) for cases and 60 (SD 74) for non-cases ( $t=3.63$ ,  $p < .005$ ). Thus negative life events (in the twelve months prior to 1983 assessment) were higher for cases than non-cases using any of the caseness definitions. As life events were only for the twelve months prior to 1983, such events were not necessarily related to reported episodes of depression (which could have occurred at any time over the subject's lifetime). Life events were taken as possible precipitating events and as index of distress but were not included in the discriminant function analysis as risk factors.

### Method

When predictor variables had been rated on several occasions, the original 1978 variables were used wherever possible. The 1978 PBI, trait depression, neuroticism and dependency scale data were used as predictor variables for analyses using caseness at

1983 and 1988 as the dependent variables. Self-esteem scores for both sexes had improved significantly over time, but most markedly from 1978 to 1983 (Tables 7.16 and 9.1). Hence the 1978 self-esteem score was used as the predictor variable in the equation for 1983 caseness, and the 1983 self-esteem scores were used to predict caseness at 1988.

As most subjects were single at 1978, presence of a 'core tie' and IBM scores were taken at 1983 and 1988 in the respective analyses. Analyses were undertaken using either presence or absence of a 'core tie' and then repeated using IBM scores. This strategy was employed because 'absence of a core tie' and 'presence of a dysfunctional intimate relationship' are not necessarily equivalent risk factors for depression. A further methodological consideration was that those without partners against whom to score the IBM returned missing values for IBM scales and were thus automatically excluded from those analyses where IBM scores were entered.

#### Choice of dependent variables

Outcome variables were derived using three different diagnostic groupings. As the numbers of cases and non-cases varied for the different diagnostic groupings, the numbers of cases and non-cases is reported for each analysis on the appropriate table. The diagnostic groupings are (i) DSM-III cases (major depression plus dysthymia); (ii) RDC cases (RDC major depression (definite and probable), episodic and (definite) intermittent minor depression categories), and (iii) 'defined cases' as described in Chapter 8.

Firstly, the analysis sought to discriminate between those who were lifetime cases (irrespective of when the first episode had occurred) and those who had never been a case. Here 1983 lifetime prevalence rates for DSM-III and RDC cases were collected using data from the 1983 DIS assessment for pre-1978 and 1978-1983 rates which together constitute 1983 cases in the respective systems. Lifetime rates to 1988

utilised 1983 lifetime rates from 1983 DIS data with addition of the 1983-88 rates from the 1988 data (which corresponds with the 1983/88 rates reported in Table 9.14).

Some subjects had already experienced an episode of depression prior to the first collection of data in 1978 so that a second strategy excluded any subjects who were cases prior to 1978. Here, the same predictor variables were used to discriminate between those who had become cases since 1978 (termed 'new cases') and those who had never been a case (with those who were cases prior to 1978 being excluded). This is not as fully prospective analysis of 1978 predictors, as some current psychosocial variables included when considered more appropriate (e.g. presence of intimate partner or 'core tie' at 1978 was not considered as an important predictor variable as most subjects were likely to be single, considering their age and student status. Instead, partner status at 1983 was considered to be a more relevant predictor variable).

#### Discriminant function analysis predicting caseness at 1983 and 1988

The canonical discriminant function loadings for the predictor variables using the three diagnostic groupings are reported for (i) the sample in 1983 (Table 10.1) and in 1988 (Table 10.5), and (ii) with those subjects who were cases (using the appropriate caseness definition) prior to the commencement of the study excluded at 1983 (Table 10.2) and 1988 (Table 10.6). Analyses for the total group used gender as one of the predictor variables, but were repeated separately for females at 1983 and 1988 (Tables 10.3 and 10.7) and for males, predicting caseness at 1983 and 1988 (Tables 10.4 and 10.8). The smaller number of males does mean that there is low statistical power for small and medium effects but these analyses are included as they show trends for male subjects. The numbers of cases and non-cases for each analysis is included in the appropriate column.



Table 10.1      Standardised discriminant function coefficients  
for lifetime cases and noncases at 1983, using the  
three caseness definitions, for the total group (n=165)

Variable	Year	DSM-III		RDC		Defined	
		+	-	+	-	+	-
Sex		.09		.11		.12	
Maternal care	1978		-.14	.20			-.04
Paternal care	1978	.60*		.52*		.67*	
Maternal protection	1978		-.48		-.36		-.32
Paternal protection	1978	.68*		.32		.24	
Neuroticism	1978		-.18		-.03		-.12
Trait depression	1978		-.43		-.54*		-.60*
Dependency	1978		-.27		-.05		-.06
Self-esteem	1978		-.09	.07		.26	
BSRI (femininity) #	1983	.27			-.25		-.05
Core tie	1983	.08		.17		.29	
Group means	non-case	.24		.38		.35	
	case		-.78		-.52		-.80
Number cases/noncases		39/126		71/94		52/113	
Chi square	(df=10)	25.70		27.72		37.41	
Significance		<.01		<.001		<.001	

\*      p < .05  
 BSRI# = Bem Sex Role Inventory, femininity subscale  
 Self-esteem: higher scores reflect low self-esteem  
 Core tie: higher score indicates presence of intimate

Table 10.2      Standardised discriminant function coefficients  
for subjects who were 'new cases' after 1978 and  
noncases at 1983, using three caseness definitions

Variable	Year	DSM-III		RDC		Defined	
		+	-	+	-	+	-
Sex		.18		.04		.26	
Maternal care	1978	.18		.42		.08	
Paternal care	1978		-.65*	.35		.74*	
Maternal protection	1978	.54			-.36		-.40
Paternal protection	1978		-.71*	.29		.42	
Neuroticism	1978	.08		.14			-.13
Trait depression	1978	.35			-.49		-.52*
Dependency	1978	.43			-.27		-.11
Self-esteem	1978	.14		.08		.33	
BSRI (femininity) #	1983		-.16	.23			-.08
Core tie	1983	.09			-.17	.17	
Group means	non-case		-.15	.22		.22	
	case	.78			-.59		-.21
Number cases/noncases		25/126		35/129		52/165	
Chi square (df=10)		15.80 ns		14.73 ns		29.55 <.005	

\*  $p < .05$

BSRI# = Bem Sex Role Inventory, femininity subscale  
Self-esteem: higher scores reflect low self-esteem  
Core tie: higher score indicates presence of intimate

Table 10.3      Standardised Discriminant Function Coefficients  
for lifetime cases and noncases at 1983, using the  
three caseness definitions, for females (n=109)

Variable	Year	DSM-III		RDC		Defined	
		+	-	+	-	+	-
Maternal care	1978	.05		.12		.05	
Paternal care	1978	.54*		.56*		.70*	
Maternal protection	1978		-.68*		-.68*		-.53*
Paternal protection	1978	.56*		.48*		.39	
Neuroticism	1978		-.21		-.15	.10	
Trait depression	1978		-.36		-.40		-.54*
Dependency	1978	.23		.27			-.08
Self-esteem	1978	.08			-.03	.47*	
BSRI (femininity) #	1983	.39			-.14	.01	
Core tie	1983		-.01	.20		.22	
Group means	non-case	.33		.54		.48	
	case		-1.05		-.70		-1.06
Number cases/non-cases		26/83		48/61		35/74	
Chi square	(df=9)	29.29		31.95		41.02	
Significance		<.005		<.001		<.001	

\* p < .05

BSRI# = Bem Sex Role Inventory, femininity subscale

Self-esteem: higher scores reflect low self-esteem

Core tie: higher score indicates presence of intimate

Table 10.4      Standardised discriminant function coefficients  
for lifetime cases and noncases at 1983, using the  
three caseness definitions, for males (n=56)

Variable	Year	DSM-III		RDC		Defined	
		+	-	+	-	+	-
Maternal care	1978		-.36		-.50	.20	
Paternal care	1978	.53			-.11		-.36
Maternal protection	1978	.26			-.72	.32	
Paternal protection	1978	.91		.10			-.13
Neuroticism	1978		-.32		-.24	.26	
Trait depression	1978		-.01	.64		.48	
Dependency	1978		-.70	.48		.42	
Self-esteem	1978	.26		.29		.20	
BSRI (femininity) #	1983		-.30	.53		.42	
Core tie	1983	.43			-.03		-.41
Group means	non-case	.24			-.26		-.27
	case		-.80	.40		.68	
Number cases		13/43		23/33		17/39	
Chi square (df=9)		8.26		4.85		7.67	
Significance		ns		ns		ns	

\*      p < .05  
 BSRI# = Bem Sex Role Inventory, femininity subscale  
 Self-esteem: higher scores reflect low self-esteem  
 Core tie: higher score indicates presence of intimate

Table 10.5 Standardised discriminant function coefficients for lifetime cases and noncases at 1988, using the three caseness definitions, for the total group (n=161)

Variable	Year	DSM-III		RDC		Defined	
		+	-	+	-	+	-
Sex			-.06		-.15	.16	
Maternal care	1978	.13			-.05	.07	
Paternal care	1978		-.45		-.42*	.52*	
Maternal protection	1978	.60*		.42			-.31
Paternal protection	1978		-.65*		-.34	.31	
Neuroticism	1978	.31			-.16		-.14
Trait depression	1978	.14			-.06		-.46*
Dependency	1978	.04		.30			-.01
Self-esteem	1983	.41		.47*			-.17
BSRI (femininity) #	1983	.01		.30		.06	
Core tie	1988		-.38		-.24	.42*	
3 or more children	1988	.10		.21			-.14
Group means	non-case		-.28		-.56	.39	
	case	.64		.56			-.85
Number cases/non-cases		50/111		80/81		51/110	
Chi square (df=11)		24.65		39.75		41.84	
Significance		<.05		<.001		<.001	

\* p < .05  
 BSRI# = Bem Sex Role Inventory, femininity subscale  
 Self-esteem: higher scores reflect low self-esteem  
 Core tie: higher score indicates presence of intimate  
 3 children: higher score indicates presence

Table 10.6 Standardised discriminant function coefficients for subjects who became cases after 1978 and those who were noncases at 1988, using three caseness definitions

Variable	Year	DSM-III		RDC		Defined	
		+	-	+	-	+	-
Sex		.15			-.13	.25	
Maternal care	1978	.13			-.08	.14	
Paternal care	1978		-.44		-.27	.61*	
Maternal protection	1978	.61		.44			-.37
Paternal protection	1978		-.74*		-.46	.49*	
Neuroticism	1978	.28			-.09		-.15
Dependency	1978	.15		.21			-.03
Trait depression	1978	.07		.32			-.41
Self-esteem	1983	.32		.48*			-.08
BSRI (femininity) #	1983	.19		.32			-.05
Core tie	1988		-.51		-.38	.44*	
3 or more children	1988	.29			.21		-.10
Group means	non-case case		-.20		-.39	.29	-1.06
Number cases/non-cases		36/111		44/81		27/110	
Chi square (df=11)		24.65		27.03		34.61	
Significance		<.05		<.005		<.001	

\* p < .05

BSRI # = Bem Sex Role Inventory, femininity subscale  
 Self-esteem: higher scores reflect low self-esteem  
 Core tie: higher score indicates presence of intimate  
 3 children: higher score indicates presence

Table 10.7      Standardised discriminant function coefficients  
for lifetime cases and noncases at 1988, using the  
three caseness definitions, for females (n=108)

Variable	Year	DSM-III		RDC		Defined	
		+	-	+	-	+	-
Maternal care	1978	.18		.03		.20	
Paternal care	1978		-.42		-.55*	.61*	
Maternal protection	1978	.93*		.58*			-.52*
Paternal protection	1978		-.67		-.45	.46*	
Neuroticism	1978	.42		.27			-.10
Trait depression	1978	.05		.29			-.36
Dependency	1978		-.01		-.07	.05	
Self-esteem	1983	.10		.26		.20	
BSRI (femininity) #	1983		-.18	.25		.14	
Core tie	1988		-.22		-.14	.35	
3 or more children	1988		-.02	.20			-.04
Group means	non-case case		-.69		-.70	.51	-1.05
Number cases/non-cases		35/73		58/50		36/72	
Chi square	(df=10)	24.11		34.29		41.91	
Significance		<.05		<.001		<.001	

\* p < .05

BSRI # = Bem Sex Role Inventory, femininity subscale  
Self-esteem: higher scores reflect low self-esteem  
Core tie: higher score indicates presence of intimate  
3 children: higher score indicates presence

Table 10.8      Standardised discriminant function coefficients  
for lifetime cases and noncases at 1988, using the  
three caseness definitions, for males (n=53)

Variable	Year	DSM-III		RDC		Defined	
		+	-	+	-	+	-
Maternal care	1978		-.12		-.34	.11	
Paternal care	1978		-.60		-.16		-.44
Maternal protection	1978		-.33	.01		.20	
Paternal protection	1978		-.88*		-.65		-.56
Neuroticism	1978	.28			-.11	.37	
Trait depression	1978	.07		.68		.41	
Dependency	1978	.49		.25		.52	
Self-esteem	1983	.82*		.95*		.80*	
BSRI (femininity) #	1983		-.18	.50		.43	
Core tie	1988		-.68*		-.42		-.70*
3 or more children	1988	.48		.21		.58*	
Group means	non-case		-.43		-.47		-.53
	case	1.12		.62		1.37	
Number cases/non-cases		15/38		22/31		15/38	
Chi square		17.40		11.84		21.29	
Significance		ns		ns		<.05	

\*      p < .05  
 BSRI# = Bem Sex Role Inventory, femininity subscale  
 Self-esteem: higher scores reflect low self-esteem  
 Core tie: higher score indicates presence of intimate  
 3 children: higher score indicates presence



The analyses were undertaken with and without IBM scores. As the IBM scores made little contribution to most of the results, the Tables 10.1 to 10.8 report data without the IBM scores. However, the complete results reporting data with and without IBM scores are available in Appendix III.

In determining whether the predictor variables are able to discriminate between cases and non-cases, the degree to which Wilks' lambda departs from unity and the significance of the associated chi square are first examined, then the levels of significance for individual variables are considered to determine the contribution of individual predictor variables, with those variables with weights for canonical coefficients greater than 0.30 being taken as important discriminators (reported in bold print in Tables 10.1 to 10.8). The significance level for individual predictor variables is influenced by the number and characteristics of the specific variables entered, and the number of subjects used in the analysis. As the discriminant function aims to produce the best linear combination of variables to discriminate between cases and non-cases, so the sign of the weights (positive or negative) can vary from one analysis to the next but examination of group means (given in Tables 10.1 to 10.8) aids consideration as to whether caseness is associated with positive or negative values in each context. One can also examine the univariate analyses to contrast how high or low values of the variables function independently as risk factors, in comparison to how they function in the linear combination defined by the discriminant analysis. Tables 10.10 to 10.12 report the data for cases and non-cases at 1988 (using the three caseness definitions) to allow the pattern of results to be observed (see pages 241-243).

### **Importance of the risk factors in predicting caseness**

#### **i) Discrimination between those who were cases at any time and those who were non-cases at the 1983 and 1988 assessments**

Table 10.1 reports data from analyses using 1983 diagnostic groupings to define caseness. The same pattern of results is seen for each grouping. High maternal protection and trait depression scores were associated with caseness and high scores for paternal care and protection are associated with non-caseness. The chi squares for all analyses are statistically significant. There is a 70% to 74% correct classification rate, with greatest discrimination between for the 'defined' case and non-case system.

Table 10.5 reports data from analyses using 1988 diagnostic groupings to define caseness. The PBI variables perform in the same fashion as for the 1983 analyses and are of equal importance for all diagnostic groupings. For the DSM-III grouping, high neuroticism and low self-esteem is associated with caseness and 'presence of a core tie' with non-caseness. High dependency, low self-esteem and high femininity scores are associated with RDC caseness, while for 'defined' caseness, high trait depression is additionally associated with caseness and 'presence of a core tie' with non-caseness. The chi squares for all analyses are statistically significant. Correct classification rates range from 70% to 73%, with greatest discrimination for the 'defined' case system.

#### **ii) Discrimination between those who were cases after 1978 and those who were non-cases**

Table 10.2 reports results from analyses discriminating between those who had become cases after 1978 (up to 1983) and those who had never been cases (by 1983). High maternal protection and trait depression scores are associated with caseness, and high paternal care scores associated with non-caseness for all three groupings. High paternal protection scores are associated with non-caseness for DSM-III and 'defined' case groupings. High dependency scores are associated with DSM-III caseness and low

self-esteem scores with 'defined' caseness. Only the chi square for discrimination between 'defined' cases and non-cases reaches statistical significance. The correct classification rates range from 69% for DSM-III caseness to 74% for 'defined' caseness.

Table 10.5 reports results from analyses discriminating between those who had become cases after 1978 (up to 1988) and those who had never been cases (by 1988). The same patterns noted in previous analyses are repeated (except that paternal care is less important as a discriminator between RDC cases and non-cases). Low self-esteem is associated with caseness for DSM-III and RDC groupings, high trait depression is associated with caseness for RDC and 'defined' cases and high femininity scores are also associated with RDC caseness. 'Presence of a core tie' is associated with non-caseness for all three caseness definitions. The chi squares for analyses discriminating between RDC and 'defined' cases are statistically significant and correct classification rates are from 66% for DSM-III caseness to 72% for 'defined' caseness.

Gender makes no important contribution to any of the analyses reported in Tables 10.1, 10.2, 10.5 or 10.6. There are some variables that are of more relevance when the sexes are examined individually and separate analysis of data for each sex will now be reported.

iii) Discrimination between those who had been cases at any time and those who had never been cases, for females only

Table 10.3 reports results from analyses using 1983 caseness definitions as the dependent variables. High maternal protection and trait depression scores are associated with caseness, high paternal care and protection scores are associated with non-caseness. High femininity scores are associated with non-cases for the DSM-III grouping and low self-esteem for the 'defined' cases. The chi squares for all three analyses reach statistical significance and correct classification rates range from 75% for RDC caseness to 80% for 'defined' caseness.

Table 10.7 reports results from analyses using 1988 caseness definitions as the dependent variables. Again, the same consistent pattern for PBI scores is seen. High neuroticism scores are associated with DSM-III cases and high trait depression scores with 'defined' cases. The chi squares for all analyses reach statistical significance and the correct classification rates range from 70% for DSM-III and RDC caseness to 79% for 'defined' caseness.

iv) Discrimination between those who had been cases at any time and those who had never been cases, for males only

Table 10.4 reports results from analyses using 1983 caseness definitions as the dependent variables. None of the chi squares for the discrimination analyses reach statistical significance and the patterns of results are not as consistent as in the analyses reported thus far. The correct classification rates vary from 61% for RDC caseness to 66% for 'defined' caseness.

Table 10.8 reports results from analyses using 1988 caseness definitions as the dependent variables. Only the chi square for the 'defined' case grouping reach statistical significance, however there are some consistent trends. High dependency scores are associated with caseness and high paternal protection scores with non-caseness for all three caseness definitions. For defined cases, there is an association with high scores on neuroticism, dependency, trait depression and femininity scales and low self-esteem scores, and the 'presence of three or more children'. For non-cases, there is an association with high scores on paternal care and protection and 'presence of a core tie'.

Further analyses were done for females alone, introducing the variable 'home duties only' (Table 10.9), which Brown had found to be an important risk factor for working-class women, as discussed earlier. Once again, the PBI scores weight highly

and the overall patterns of results are similar. Home duties weight significantly for RDC cases only.

Tables 10.10 to 10.12 report the univariate analyses for the three lifetime (to 1988) caseness definitions to allow examination of the variables used in the discriminant function analyses. There are consistent differences between scores for cases and non-cases with cases reporting higher scores for neuroticism, trait depression and maternal protection and reporting lower scores for self-esteem, maternal (RDC and 'defined' cases only) and paternal care, while RDC cases also report lower IBM scores for care from partner.

Table 10.9 Discriminant Function Analysis for lifetime cases and noncases at 1988, using three caseness definitions for females only, with 'home duties' added

Variable	Year	Standardised Discriminant Function Coefficients		
		'Defined' Cases	RDC Cases	DSM-III Cases
Maternal care	1978	.20	-.00	.16
Paternal care	1978	.61*	-.51*	-.41
Maternal protection	1978	-.52*	.56*	.93*
Paternal protection	1978	.46*	-.50*	-.69*
Neuroticism	1978	-.10	.25	.42
Trait depression	1978	-.36	.33	.07
Dependency	1978	.05	-.08	-.01
Self-esteem	1983	.20	.28	.10
BSRI (femininity)	1983	.14	.28	-.17
Core tie	1983	.35	-.20	-.24
3 or more children	1988	-.04	.16	-.03
Home duties only	1988	-.00	.31	.11
Group means	non-case	.51	-.73	-.36
	case	-1.05	.63	.78
Wilks' Lambda		.648	.684	.777
Chi square		41.69	36.48	24.20
Significance		<.001	<.001	<.05
Sensitivity		75.0%	72.4%	68.6%
Specificity		80.6%	82.0%	76.7%
Correct classification		78.7%	76.9%	74.1%

\* p < .05

# BSRI = Bem Sex Role Inventory, femininity subscale  
 Self-esteem: higher scores reflect low self-esteem  
 3 children: higher score indicates presence  
 Core tie : higher score indicates presence of intimate  
 IBM care : higher score indicates low care  
 IBM control: higher score indicates high control

Table 10.10 Comparison of scores used in discriminant function analysis for DSM-III lifetime cases and non-cases to 1988

Measure	Year assessed	Case Mean (SD)	Non-case Mean (SD)	t-test
<u>Personality measures</u>				
Trait depression	1978	34.53 (11.97)	29.38 (11.27)	-2.55*
Dependency	1978	53.73 (10.01)	52.19 ( 9.49)	-0.91
Neuroticism	1978	10.18 ( 4.99)	8.39 ( 4.50)	-2.17*
Self-esteem <sup>#</sup>	1978	1.86 ( 1.41)	1.42 ( 1.45)	-1.80
Self-esteem	1983	1.44 ( 1.67)	0.78 ( 1.08)	-2.56*
Femininity	1983	4.74 ( 0.48)	4.78 ( 5.00)	0.46
<u>PBI scores</u>				
Maternal care	1978	24.96 ( 7.70)	26.88 ( 6.38)	1.54
Paternal care	1978	19.38 ( 9.15)	22.93 ( 8.06)	2.36*
Maternal protection	1978	16.64 ( 7.25)	13.65 ( 7.42)	-2.41*
Paternal protection	1978	13.24 ( 7.54)	13.28 ( 7.22)	0.03
<u>IBM scores</u>				
Partner care	1983	29.15 ( 5.86)	31.14 ( 5.17)	1.84
Partner control	1983	6.97 ( 5.47)	6.53 ( 4.80)	-0.44
Self-esteem <sup>#</sup> : high score denotes low self-esteem				
* p <.05    ** p <.01    *** p <.001				

Table 10.11 Comparison of scores used in discriminant function analysis for RDC lifetime cases and non-cases to 1988

Measure	Year assessed	Case Mean (SD)	Non-case Mean (SD)	t-test
<u>Personality measures</u>				
Trait depression	1978	34.35 (12.18)	27.63 (10.21)	-3.77***
Dependency	1978	53.87 ( 9.30)	51.48 ( 9.90)	-1.57
Neuroticism	1978	10.05 ( 4.84)	7.85 ( 4.35)	-3.03***
Self-esteem <sup>#</sup>	1978	1.75 ( 1.39)	1.37 ( 1.49)	-1.67
Self-esteem	1983	1.40 ( 1.58)	0.57 ( 0.83)	-4.16***
Femininity	1983	4.80 ( 0.47)	4.78 ( 5.21)	-0.92
<u>PBI scores</u>				
Maternal care	1978	24.83 ( 7.19)	27.72 ( 6.21)	2.73**
Paternal care	1978	19.86 ( 9.31)	23.78 ( 7.26)	2.98**
Maternal protection	1978	16.50 ( 7.28)	12.68 ( 7.21)	-3.35***
Paternal protection	1978	13.98 ( 8.11)	12.56 ( 6.37)	-1.23
<u>IBM scores</u>				
Partner care	1983	29.25 ( 5.90)	31.78 ( 4.67)	2.72**
Partner control	1983	7.27 ( 5.49)	6.09 ( 4.46)	-1.35
Self-esteem <sup>#</sup> : high score denotes low self-esteem				
* p <.05    ** p <.01    *** p <.001				



Table 10.12 Comparison of scores used in discriminant function analysis for 'defined' lifetime cases and non-cases to 1988

Measure	Year assessed	Case Mean (SD)	Non-case Mean (SD)	t-test
<u>Personality measures</u>				
Trait depression	1978	37.10 (12.31)	28.16 (10.30)	-4.47***
Dependency	1978	53.30 ( 9.94)	52.38 ( 9.55)	-0.55
Neuroticism	1978	10.49 ( 4.55)	8.23 ( 4.64)	-2.92***
Self-esteem	1978	1.82 ( 1.49)	1.44 ( 1.42)	-1.56
Self-esteem	1983	1.41 ( 1.58)	0.79 ( 1.14)	-2.53*
Femininity	1983	4.69 ( 0.47)	4.81 ( 0.50)	-1.35
<u>PBI scores</u>				
Maternal care	1978	24.53 ( 7.66)	27.55 ( 6.06)	3.30***
Paternal care	1978	17.66 ( 8.67)	23.91 ( 7.98)	4.30***
Maternal protection	1978	17.65 ( 7.56)	13.16 ( 7.02)	-3.59***
Paternal protection	1978	14.52 ( 8.26)	12.65 ( 6.95)	-1.39
<u>IBM scores</u>				
Partner care	1983	29.53 ( 4.72)	30.94 ( 5.66)	1.45
Partner control	1983	7.39 ( 5.56)	6.39 ( 4.78)	-0.96
Self-esteem <sup>#</sup> : high score denotes low self-esteem				
* p <.05    ** p <.01    *** p <.001				

## **Discussion**

### **The relationship shown for the putative risk factors and depression**

This discussion is not intended as an exhaustive overview of an extensive body of literature and will deal particularly with vulnerability factors as quantified by the self-report measures used in this study. It is also important to recognise that risk factors to mild and moderate levels of depression (which usually implies non-psychotic, non-melancholic unipolar depression) may differ very much from risk factors to severe depression (which implies psychotic or melancholic depression associated with unipolar or bipolar disorders). As noted earlier, most of the research into personality and vulnerability factors has concentrated on the mildly to moderately depressed groups and Bebbington (1988) has summarised a position that would be commonly held: " My own speculation is that moderate and severe depression are different entities, that the age effect in the former arises because of social influences, and that ageing exerts a (probably biological) effect on severe depressions. This makes the age effect in severe depressions a default option, open to reversal, but only by social factors of the most adverse character". Thus in a group of young adults in a non-clinical cohort, one would expect to be encountering mild to moderate levels of depression where the impact of psychosocial risk factors would be most evident.

Different factors may determine onset, outcome and chronicity of depressive episodes. The presence of chronic depression itself may also influence scores in measures that attempt to quantify risk factors, either as a direct state-dependent effect or indirectly (e.g. by such effects as decreasing self-esteem or unemployment). By 1988, most depressed members of the cohort had experienced no more than three discrete depressive episodes and were then in their mid-thirties, consequently, the presence of chronic depression is unlikely to be a major confounding factor for this cohort.

### Neuroticism as a risk factor

Neuroticism, as measured by the Eysenck Personality Inventory (Eysenck & Eysenck, 1964) has been held to be a key personality risk factor to onset of depression as it "encompasses vulnerability to breakdown under stress and a proneness to anxiety and emotional instability" (Katz & McGuffin, 1987). In the present cohort, cases report higher scores for neuroticism in the univariate analyses but high trait neuroticism did not make a consistently significant contribution in discriminating between cases and non-cases, but had some relevance as a discriminator for DSM-III cases.

High neuroticism scores have also been associated with unresolved depression (Weissman et al, 1978; Boyce and Parker, 1985) and are held to predict chronicity (Hirschfield et al, 1986) but, in both these studies, subjects were already depressed when the neuroticism scores were first measured so that it was not clear whether high scores reflected a state and/or trait effect. The finding of moderate levels of consistency in neuroticism scores over ten years in the teachers' cohort (as reported in Chapter 9), despite relatively high rates of depressive disorders, would argue that the EPI neuroticism scale is a trait measure, albeit with some state sensitivity during actual depressive episodes.

### Dependency as a risk factor

A personality trait of dependency has also been considered to dispose to depression (Birtchnell, 1984). Birtchnell stated that Blatt's construct of dependency used in the dependency scale (Blatt et al, 1975) was based on the concept of anaclitic depression, which was related to "feelings of vulnerability to feelings of deprivation, intense need for support and gratification from others and difficulty in managing anger for fear of destroying the object and the gratification it provides". Blatt's dependency construct may well be contaminated by state depression effects but was judged to be the best available at the commencement of the present study (in 1978). Birtchnell's (1984) review of the relationship between dependence and depression makes the point that

dependence is a personality trait which carries an increased predisposition to depression but at the same time, depressed individuals are more likely to display dependent behaviour.

In the univariate analyses, females score higher than the males on the dependency (Chapter 7) but there were no differences between dependency scores for cases and non-cases in the univariate analyses reported here (Tables 10.10 to 10.12). High dependency scores are weighted (and associated with caseness) in discriminant analyses for males only (for 'defined' caseness at 1983 and all caseness definitions at 1988). Against expectation, dependency appears more important as a risk factor for males than for females in this cohort. It is possible that male teachers as a group may differ from males in other professions (e.g. engineering, economics) including being more dependent or because the nature of their work involves interpersonal demands of the sort that are more commonly encountered in female-orientated careers (e.g. nursing). This issue will be pursued in the next chapter.

The notion of interpersonal sensitivity, measured by the Interpersonal Sensitivity Measure or IPSM (Boyce & Parker, 1989), may be a more discriminating risk factor than dependency but the measure was not available at the commencement of the study. IPSM variables such as 'interpersonal awareness' and 'timidity' have been reported as important vulnerability factors to depression in young women (Boyce et al, 1990) and the IPSM will be rated by the cohort at subsequent reviews.

#### Low self-esteem as a risk factor

Low self-esteem has been considered as a predisposing factor to depression but also as part of the process of, or a sequel to, a depressive episode (Ingham et al, 1986; Brown et al, 1986; Robson, 1988). Robson (1988) has criticised Rosenberg's (1965) scale (the self-esteem measure used in this study) as being too global in concept, quantifying such constructs as personal worth, appearance and social competence, but

not reflecting the "richer intuitive clinical idea". However, once again, the commencement of the present study pre-dated these criticisms. The Rosenberg measure was also used by Ingham's group to gauge self-esteem in a group of Edinburgh women. They found an association between maintenance of intimate relationships and self-esteem and concluded that changes in self-esteem were not a consequence of illness or mood changes. They postulated that the presence of a close confidant enhances self-esteem but those who lack self-esteem will have difficulties maintaining a close relationship. They considered self-esteem to be a trait measure but suggested that the presence of small children may have an effect on self-esteem, for women who develop a diagnosable depression. Their proposed relationship between self-esteem and presence of a confiding partner leaves one unclear whether they are proposing that the presence of a confidant engenders self-esteem or that those with high self-esteem will be more able to initiate and maintain a confiding relationship.

In the univariate analyses of results for the teacher cohort, the difference in means between cases and non-cases is statistically significant for the 1983, but not for the 1978 self-esteem scores. Self-esteem scores are an important discriminator between cases and non-cases for the 'defined' case grouping (for females and the group as a whole) at 1983 (where the 1978 self-esteem scores were used). In 1988, when the 1983 self-esteem scores are used as depression predictors, self-esteem is an important discriminator between cases and non-cases for DSM-III and RDC caseness definitions (for males and the whole group) and to a lesser extent for 'defined' cases (for males), with high scores (reflecting low self-esteem) associated with caseness.

Cox (1990) made an interesting proposition concerning self-esteem in young women following his obstetric experience in Africa. He noted that, in that culture, the event of childbirth is accompanied by 'rites of passage' or rituals which are lacking in our society. He proposed that the absence of these rituals may relate to onset of depression by "lowering the mother's self-esteem, causing uncertainty about the

availability of social support, increasing the likelihood of physical fatigue, as well as by stressing the relationship with her husband. Furthermore, as the lack of ritual structure post-partum represents the ambivalence about the status to be attached to mothering, this uncertainty exacerbates her role conflict and increases the threat to her self-esteem". In the current study, self-esteem scores did change over time, indicating either a state effect or an actual change (which was particularly evident in the first few years of work life) due to maturational processes and social role changes within group members. Cox's observations may go some way towards explaining the trend for females to report more episodes of depression over the 1983-88 period (noted in Chapter 8), when many of the females were having babies, but there was not an observed lowering of self-esteem in females, rather a greater improvement of self-esteem in males. However, Cox's comments on lack of ritual in contemporary Western society could apply equally to men and build to Murphy's (1986) comments on changing roles for men (i.e. that men are being expected to take a more nurturant role in their interpersonal relationships) leading to increased vulnerability to depression for men which are discussed further in the next chapter.

#### Perception of poor parenting as a risk factor

A causal relationship has also been postulated between anomalous parenting (as measures by the Parental Bonding Instrument or PBI), self-esteem and non-melancholic depression. PBI scores have been demonstrated as stable over time and free of mood state effects (Parker, 1989; Gotlib et al, 1988). Parker (1988) has stated that "care from a parent appears to have a key influence in setting the child's inherent self-esteem, or the degree to which it intrinsically values itself. A lack of care (whether by indifference, rejection, critical comments or hostility) promotes a low self-esteem and a diathesis to adult depression". A further finding, that the likelihood of depression following loss of parent in childhood is influenced both by the child's perception of the quality of care received before and following the crisis and the quality of any later marital relationship (Parker & Hadzi-Pavlovic, 1984), reflects the interaction between

deficits in parental environment, the perception of the child's needs by parents or carers 'in loco parentis' and the later acquisition of self-esteem.

PBI scores were completed by the cohort before many of the group had experienced depressive episodes, so that causal inferences can be made. Parker's view is borne out by the present findings, where PBI scores, most notably low paternal scores (care and protection) and high maternal protection scores, are consistent discriminators of cases and non-cases, no matter which classificatory system is used. This pattern of findings is still present when analyses included only those who had become cases after the first PBI measures had been completed. In the analyses, perception of high maternal protection is consistently associated with depressive caseness, while perception of high paternal protection, along with high paternal care is associated with non-caseness. On inspection of the univariate analyses, paternal protection scores do not significantly differ between cases and non-cases (unlike the other paternal and both maternal PBI scales). It is possible that there are different perceptions of the construct of protection related to the sex of the parent and in this cohort, the constructs of paternal care and protection may have been viewed more similarly than maternal care and protection. The two patterns of PBI scores found in the discriminant analyses suggest either (i) very high maternal protection could be associated with lower (but 'good enough') paternal care and protection scores and/or (ii) average ('good enough') maternal protection scores associated with very low paternal scores. The interactions between the variables cannot be further clarified by such a multivariate analysis.

There are also some sex-related effects in terms of ranking of canonical coefficients in the discriminant function analyses, with females ranking maternal PBI scores more highly and males ranking paternal scores more highly. This would indicate that while high PBI scores attributed to either parent may be important, the effect of the same-sex parent may be the more important in vulnerability to depression, as noted in earlier PBI research (Parker, 1983b).

The effects of perceived lack in the parental environment and poor self-esteem are common to both sexes and, in this group, these variables are far more important than the effect of gender itself in determining possible onset of depression. As noted, the relationship between high maternal protection and low paternal scores for care and protection in the discriminant function analyses may reflect either maternal compensatory over-concern for an emotionally distant father (either through depression, illness or indifference) or a father who is deemed to be distant and ineffective because of his inability both to father and to intervene with an intrusive and over-protective mother. It is beyond the scope of this study to determine which factors in the parental environment could determine the constellation found by the discriminant function results but possible determinants include (i) a chronically poor marital relationship, (ii) an absent father and harried mother, or (iii) the presence of depression in either (or both) parents.

A causal link is hypothesised, with low parental care and high maternal protection leading to low self-esteem and high trait depression. It should be noted that the results of the discriminant function in this group cannot be generalised to other groups, although these findings are in broad accordance with the other research cited. Further studies, particularly of a longitudinal nature, starting prior to birth, would be necessary to evaluate these hypotheses adequately.

#### Feminine sex role characteristics as a risk factor

Bem (1974) developed her Sex Role Inventory (or BSRI) to measure sex role characteristics, with the hypothesis that males reporting high masculinity scores (and low femininity scores) or females reporting high femininity scores (and low masculinity scores) were seen as operating within sex-role stereotypes which were more inflexible and psychologically constricting. Those who scored highly for both masculinity and femininity items were termed 'androgynous' and hypothesised to be more flexible in



outlook and less prone to psychological distress. The extent to which similar scores for the sexes reflect intrinsic personality characteristics and/or the outcome of socialization processes cannot be estimated, but assessing the relevance of sex role inventory scores to groups demonstrating sex differences in depressive disorder might provide useful information about mediating mechanisms. The BSRI includes an additional social desirability sub-scale for which there were no sex differences (Table 7.16), with both sexes returning high scores, consistent with their self-esteem scores.

In this cohort, the scores on the masculinity subscale show no sex differences (with females rating highly on the masculinity scale, defined by such descriptors as 'self-reliant', 'ambitious' and 'aggressive'). The femininity scale is scored more highly by females and show a statistically significant sex difference at ten year follow-up in 1988 (Table 7.16). Despite this, the femininity construct is more highly related to depression in males, so that males who affirm such descriptors as 'sympathetic', 'understanding' and 'gullible' are more likely to be classified as depressive cases. These findings are similar to those seen for the dependency variable and fit with Murphy's (1986) suggestion that males are now being expected to assume more active roles as partners, rather than simply be bread-winners, but this change to a more nurturant role may increase their vulnerability to depression in the mild to moderate range, although she also conjectured that there may be a simultaneous decrease in more serious psychiatric morbidity (suicide, alcoholism), which remains to be seen.

In the discriminant function results, high femininity scores were associated with caseness for males, an important finding suggesting that sex role may be as important to study as gender sex in any community with a female preponderance in depression rates.

#### High trait depression as a risk factor

A nine-year study of over 8,000 adults (Kaplan et al, 1978) in California, USA, incorporated the Center for Epidemiological Studies Depression Scale or CES-D

(Radloff, 1977) to measure depressive state. They found that the base-line level of depressive symptoms was the most powerful predictor of high symptom levels nine years later, despite the fact that 78% of the sample did not report high symptoms on either occasion. Despite the observation (reviewed in Chapter 3) that females tend to rate more highly on depressive symptom measures, Kaplan's group concluded that gender itself had no predictive value and that the increased long-term risk of high levels of depressive symptoms was due to gender-related characteristics rather than gender itself. Scores for the state measure used here (Wilson, 1979) are poorly correlated over time (see Chapter 9), and, in that respect, consistent with what one would expect of a state measure. Despite the possibility that the 1978 state depression scores may have predicted depressive caseness, this variable was not used as a predictor in the discriminant analysis because of the potential confounding effect of using a measure of depressive symptoms to predict lifetime cases of depression (where caseness was largely based on reports of depressive symptoms).

The mean scores for the trait depression measure (Costello and Comrey, 1967) are consistently higher for cases (compared to non-cases) using any of the caseness definitions but there is the possibility that the instrument is simply measuring depressive state. However, as previously noted (in Chapter 9), the trait depression scores are consistent over ten years, unlike the state depression measure (Wilson, 1979). The items in the trait depression measure are intended to reflect a long-term negative cognitive set, which the authors conceptualised as "a person's tendency to experience a depressed mood" and are relatively free of sex-role typed factors, with more potential overlap with self-esteem responses and locus of control issues than depressive symptoms. This variable is weighted highly and consistently in the discriminant function analyses, but less so than PBI scores. While there is a seeming tautology in depression (trait measure) predicting depression (actual occurrence), the finding supports the Costello-Comrey measure as achieving its objective and suggests a variable defining subjects at high risk.

### Absence of a 'core tie' as a risk factor

When considering comparisons between cases and non-cases using univariate analyses at 1988, there is a statistically significant association between the presence of a 'core tie' (or intimate partner who was a confidant) and depressive non-caseness. This effect is only apparent for 'defined' caseness at 1983 but for all caseness definitions at 1988. Differences between cases and non-cases in terms of perceived adequacy of social support in times of stress have been noted earlier and are present for all case definitions but more highly significant for 'defined' case criteria and support the view that the presence of a 'core tie' provides an important source of social support, consistent with the findings of Brown and Harris (1978) and Henderson's group (1981). The inclusion of IBM scores (which measure functional aspects of an intimate relationship) does not add substantially to the discriminant analyses (see Appendix III), other than for males, where the 'presence of three or more children' and high IBM 'control' scores were also associated with caseness. IBM scores are shown to have some relevance in discriminating cases from non-cases at 1988, but there are problems in that depressive cases are more likely not to have a partner, and are therefore dropped from the analyses using the IBM scores, so that structural and functional aspects could not be examined in the same analyses. The IBM scores are of greater relevance at 1988, which is to be expected, as many of the group were still courting or newly married at the time of the 1983 assessment. The PBI generally assumes higher canonical coefficient weights at this stage over the IBM but it is envisaged that over time, the IBM scores may gain ascendancy. By 1988, most of the subjects had still been married less than a decade and the patterns in their relationships, including the effects of small children, are still evolving.

### Female gender as a risk factor

Lastly, and most importantly, female gender is not identified in these analyses as a significant risk factor for depressive caseness. Any association between gender and

depressive caseness may be mediated through such factors as feminine sex-role, perceived lack of care from parents (in past) and partner (in present), and leaving full-time work to stay home with young children but are not clearly related to "femaleness" in the discriminant function analysis.

### Conclusions

High trait depression and neuroticism, low self-esteem, perception both of low paternal care and high maternal protection are all identified as risk factors to depression for both sexes. All results for risk factors are as hypothesised (Hypothesis 4).

The presence of a 'core tie' is commoner in non-cases than cases in the univariate analyses. In the discriminant function analyses, the presence of a core tie was associated with depressive non-caseness for the 1988 caseness definitions.

For males, perception of low paternal protection and high control from partner are important risk factors, also, in some caseness definitions, the presence of 3 or more children at home.

It was hypothesised (Hypothesis 3) that high interpersonal dependency, feminine sex role stereotype and exposure to small children at home would be particular risk factors for females. This has not been shown, and if anything, these variables constitute a greater risk factor for males in this group.

These findings are in keeping with the hypothesised causal links between the perceived parental environment and self-esteem, then affecting later depressive experience.

Results of discriminant function analyses cannot be used to generalise beyond this group but are appropriate to issues such as changing social expectations raised for contemporary young adults.

Of the three caseness definitions, the 'defined' case definition (which used DIS-generated categories, with additional operationalised help-seeking and impairment criteria) proved to be the most effective in discriminating between cases and non-cases. The DSM-III case definition proved to be the least effective. Overall, the three definitions are similar in classification rates by risk factor predictors and the same risk factors are relevant for discrimination of cases and non-cases.

## CHAPTER ELEVEN

### DISCUSSION OF FINDINGS

#### CONTENTS

##### **Issues pertaining to the cohort**

How representative is this group of the general population?  
 Did those who volunteered for the study differ from their peers?  
 What were the special characteristics of the group?  
 Were there particular issues related to teaching as a career?  
 Is there evidence that the males and females within the group differ from each other or expected social norms?  
 Are there sex differences in recall?

##### **Examination of sex differences**

Sociodemographic variables  
 Are the lack of sex differences due to a Type II error?  
 Were there sex differences in perceptions of social support networks?  
 How do the rates for DIS-derived depression categories compare with other studies?  
 How accurate are lifetime prevalence rates?  
 Why are lifetime prevalence rates higher in younger cohorts in recent times?  
 Have there been differences in social role expectations for younger cohorts?

##### **Other issues pertaining to the study**

What is the relationship between anxiety and depression?  
 What is the relationship of life events to onset of depression?  
 How do the data from the study compare with the Bedford College model for social factors related to depression?  
 What is an appropriate definition of caseness?  
 How important was gender as a risk factor to depression for the caseness definitions used in this study?

##### **Conclusions**

##### **Postscript**

There are a number of themes to be drawn together in the discussion of this study. Firstly, there are the special characteristics and issues pertaining to this cohort. Secondly, the findings from self-report measures and the semi-structured interview will be examined with particular reference to any sex differences established. Thirdly, rates of depression from DIS-generated data will be discussed in relation to rates from other comparable studies, with an emphasis on findings for the 20 to 45 year age group and notable sex differences. Data relating to analysis of risk factors to depression will be then be examined in light of the concepts of caseness used.

While the subjects in the cohort did not all go on to teaching as a career, they will be referred to as 'teachers' in this discussion for ease of identification and comparison with other studies.

### **Issues pertaining to the cohort**

#### **How representative is this group of the general population?**

The study design required selection of a group whose male and female members had a number of important social risk factors to depression initially controlled. As a consequence, and because the selected teacher trainees were predominantly middle class, their general experience of depression may differ from that of a more broadly-based general population sample. Thus, they are not a representative sample of the general population and were never intended to be so. The strength of the study lies in the choice of a cohort whose male and female members were as similar as possible at base-line, in terms of social roles, career expectations and likely experience of depression, while later diversity in all these areas was anticipated.

Compliance was high throughout, with 97% of the sample being located ten years later. However, three males were excluded from the ten-year interview due to ill-health so that 95% of the original group were fully reassessed at the end of the ten years.

#### **Did those who volunteered for the study differ from their peers?**

In 1978, prior to the start of the study, all members of a class of trainee teachers were asked to fill in an anonymous screening measure and to provide their name and address if they were willing to proceed with the study. One might speculate that responders and refusers would differ in terms of basic interest in such a study and that females might be more willing to participate than males. As the cohort had been told that the objective to observe a normal group developmentally (with depression only being mentioned as one of the issues involved and with no emphasis on psychiatric morbidity),

there is no reason to assume a strong bias influencing acceptance. Males and females showed an equal percentage acceptance rate and there were no sex differences in scores on the self-esteem rating used as the screening measure. Romans-Clarkson et al (1988) have recently established that refusers in a community survey of psychiatric disorder in New Zealand did not differ from participants in terms of rates of attendance at psychiatric outpatients clinics, or those with hospitalized medical or psychiatric illness.

### **What were the special characteristics of the group?**

The subjects in the study are notable for their level of intelligence, degree of cooperation and accessibility at follow-up. This was facilitated by the cohort's clear commitment at the commencement which involved provision of a long-term contact address. Again, they are not comparable with a randomly selected general population group. Indeed, the hypotheses were predicated on the selection of such a socially homogeneous group which disallowed selection of a general population group, where heterogeneity would have disallowed risk factors to be largely controlled at entry. They were an interesting group of people most of whom readily shared their experiences. As the subjects were volunteers, it is assumed their high motivation favoured the collection of accurate data.

### **Were there particular issues related to teaching as a career?**

The study took place against a background of political turmoil in the teaching profession. In 1978, the group was told that teaching jobs were scarce, which generally lowered morale. As most of the group already had a university degree, about 15% found alternative employment and did not ever teach, while a minority waited months, or sometimes years after graduation before gaining their first teaching job. There was also an affirmative action policy in place, whereby women were being promoted more quickly than men, without going to a rural school (which had been the time-honoured process for rapid promotion in the public school system). There was also increasing dissatisfaction with teaching conditions and the 'hardline' policies of two N.S.W.



Education Ministers (Susskind, 1987; Lumby, 1989; Kirk, 1990). As there were high levels of depressive caseness for both sexes, such issues may have played a part. Most of these issues would have affected both sexes relatively equally but the affirmative action policy, which had a more negative effect on the males, seemed to be more of a cause for some angry comments than a source of serious depression.

Finlay-Jones (1986) has reported high levels of stress in Australian teachers. He studied over 2000 teachers in Western Australia and, using the GHQ, established that 17% showed severe psychological stress, as compared to 9% of the general population. The rate was not affected by age or sex. Thus, teachers may have high levels of morbidity, preceding or as a consequence of their profession, and males and females appeared to be similarly affected. Other studies have examined teacher stress (Kyriacou & Sutcliffe, 1977; Galloway et al, 1984), usually with a cross-sectional design in the work context, but without addressing personality and other possible psychosocial risk factors, so that it is difficult to determine how much of the reported stress is solely work-related. There have been no instruments used consistently to measure stress or working conditions for teachers and none seemed to have any particular virtues. As some of the cohort were not teaching and employed in a variety of other occupations, including home-making and part-time work, the measure selected (Renwick & Lawler, 1979) was one designed for an American survey of work practices and was not intended specifically for teachers. This was considered appropriate as, in 1983, it was not known how many of the cohort were actually still teaching. The finding that females rated their expectations of work more highly in 1983 was unexpected, but in keeping with a higher level of anticipated pleasure in their work when asked in 1978. There were no such sex differences at 1988, by which time many women were engaged in part-time work or at home with small children, while male work patterns were unchanged.

Both sexes reported that the most demoralising feature of their profession was the frustration of being unable to do their job as they saw it (i.e. to impart knowledge)

because of continuing discipline problems within the classroom. A proportion stated that in the first few years of their teaching career they had been given notoriously difficult classes when they were inexperienced, rather than a more gentle introduction. Other sources of discontent commonly cited were low wages and frustrations with a cumbersome education bureaucracy. In general, there were some particular issues related to teaching but they appear to have affected males and females equally. These work-related difficulties may have contributed to the high rates of depressive disorder in the first few years after graduation but it is unlikely that there were any sex differences in perception of such stressors. If anything and against expectation, females reported more satisfaction from their work conditions at the 1983 assessment.

**Is there evidence that the males and females within the group differ from each other or expected social norms?**

Two possible explanations of the dearth of sex differences are that the females were a particularly resilient group, or that the males were particularly vulnerable. Self-esteem improved over time for both males and females, but to a greater extent for males. Females consistently reported higher scores on the neuroticism scale and dependency sub-scales (reported in Chapter 7), but there were no sex differences shown on the other measures, most notably, the trait and state depression measures. Females were far more likely to weep, over-eat and visit friends when depressed, all of which are in keeping with expected sex-role related differences and the higher femininity scores for females on the Bem Sex Role Inventory or BSRI (Bem, 1974). All of these differences are in line with expected sex-role typed behaviours but do not necessarily indicate any decreased resilience among the females. A study of depressed university students (Funabiki et al, 1980) had previously shown that depressed females were more likely to over-eat, engage in self-deprecation and avoid large social gatherings, while seeking out personal contact with friends, all of which are completely consistent with the current findings. Two further studies (Frank et al, 1988; Young et al, 1990) of more severe depressives (i.e. patients meeting criteria for at least DSM-III major depression), also found statistically

significant sex differences only for increased appetite and weight gain in both sexes and also increased rates of somatization and expressed anger but neither study demonstrated sex differences in more "core" depressive symptoms nor in global severity of depression.

There were no sex differences in the teachers' masculinity scores on the BSRI, and it is of interest that the female teachers reported higher masculinity scores, in fact, closer to the males than the females in Bem's original sample of college graduates (Bem, 1974). Another group (Steinberg et al, 1987), investigating clinically depressed men and women, have subsequently noted that females rated equally with males on the masculinity subscale while females rated higher than males on the femininity subscale, with both sexes viewing masculine sex-role typed traits as more socially desirable. This finding and the lack of sex differences in self-esteem scores (with which the construct is said to be highly correlated) are in keeping with findings from a group of Australian university students (Antill & Cunningham 1979). This means that as a group, the females are more androgynous and possibly more psychologically resilient than groups of more highly feminine sex-role typed females, but not more so than the males in the group.

The trend for a greater increase in self-esteem for males is in keeping with findings from a longitudinal study of recent graduates (Holström et al, 1987) described in Chapter 3. Those authors found that males and females performed similarly in the years postgraduation but females experienced lower self-esteem, particularly when commencing work. The women were reported as talking more freely and the males were more likely to use alcohol.

In Chapter 7, there is a trend noted for males to consume more illicit drugs and alcohol (and the trend towards a sex difference in alcohol consumption may become more apparent over time), but there are low levels of usage in both sexes. The overall alcohol consumption for all Australian adult males<sup>+</sup> is 135 gm/week and for females<sup>+</sup>, 40

<sup>+</sup>  
in the 20 to 40 year range

gm/week (Dept. of Comm. Services & Health, 1988). Both sexes were less likely to smoke cigarettes when they started work, but both sexes smoke less than the mean rates for young Australian adults where rates were 34% for males and 31% for females, with a further 26% of males and 19% of females being ex-smokers (Dept. of Comm. Services & Health, 1990).

One may expect males who select teaching as a career and endorse "working with children" as an important motivating factor to be more nurturing than the stereotypical 'macho' male. While no attempt was made to corroborate details, there seemed to be low levels of sociopathy in the group, with the only known arrest in one male being for a drug-related crime. The males reported help-seeking patterns that were similar to the females but they were significantly more likely to engage in reckless behaviour when depressed in the first few years after graduation (before they became husbands and parents). There were many symptoms and a number of coping styles that both sexes endorsed (therefore no sex differences), but no symptoms and only one coping style (recklessness in 1983) that the males endorsed more than females. In general, the males seem to be behaving like males, but are possibly more androgynous than would be expected if derived from a more heterogeneous (and less well educated) group, but were similar to Australian male university students (Antill & Cunningham 1979).

Jenkins' study (1985) of a socially homogeneous group of civil servants in their mid-20s was noted in Chapter 3, and provides the most comparable study group. She sought to "assess the likely magnitude of the contribution of biological factors to the reported sex difference in minor psychiatric morbidity by controlling and minimizing" environmental differences. In selecting a group of university graduates in their late twenties, she too chose to examine a group of males and females closely comparable in age, marital status, educational attainments, paternal social class, domestic responsibilities and overall social supports and stresses, but did not formally assess depressive episodes.

Her men and women did not differ in terms of the prevalence of minor psychiatric morbidity and she found current impairment in about one-third of each sex, but the women were more likely to report depressive symptoms than men (39% vs 28%) and the men tended to recover from episodes of depression more quickly.

In the teachers' group, there were no sex differences in number or duration of episodes. There are several possible explanations. (i) The male teachers may be less resilient than the male civil servants in Jenkins' group, or (ii) more stressed, or (iii) the male civil servants, like the young men in Angst's study (Angst & Dobler-Mikola, 1984c) may be tending to "forget" depressive episodes more quickly than females over the ensuing twelve months, whereas the male teachers who sought help (and did not resort to alcohol) may be more likely to recall depressive episodes for longer periods. The teacher cohort has reported similar help-seeking patterns to a study of young adults in Sydney, Australia (Reynolds et al, 1979) where 13-14% of adults under 26 years of age, and 11% of males and 12% of females aged 26-45 had sought professional help for personal problems, with no sex difference in either age group.

#### **Are there sex differences in recall?**

The data have been examined in a number of ways to determine consistency over time. There is moderate to high consistency for self-report scales measuring aspects of personality and perception of earlier parenting, and fair to low consistency for reports of depressive categories, help-seeking and perception of care and control by partners over time.

As the DIS has not been validated as a measure of depressive disorders, it was judged that collaborative information should be sought where possible. Such a procedure corresponds somewhat with that adopted by Leckman and colleagues (1982) for the lifetime version of the SADS, in that they interviewed probands and normals, together

with a percentage of their spouses and first degree relatives for a corroborative report, and (for the patients) they consulted medical records to derive a 'best estimate' diagnosis. It is useful here to report the experience with the 'corroborative witnesses'. Most of the subjects were willing to nominate a witness to be interviewed and, while most of the witnesses had reasonable concerns about privacy, these were allayed with explanation. The author was impressed that both parties generally understood there to be differences between sadness and depression, and between functional and dysfunctional depressive states.

A problem that became evident was that those subjects who did not wish to provide witnesses were also likely to be those who guarded their privacy and often had poorer relationships with their families, and fewer friends, so that consistency of reporting by witnesses was likely to have been low if such reporting had been possible. Thus, consistency of reporting may be artifactually high due to the exclusion of such subjects. Some of these subjects had reported frequent and long-standing episodes and reports from other sources (e.g. time away from work, medical records) would have assisted in checking the reliability of reporting in subjects who did not wish contact with a collaborative witness.

Most subjects seemed to be reporting accurately and honestly concerning significant life events and issues deemed important to them, whether they had experienced an easy or difficult time. It seems unlikely that there was a substantial practice effect operating as most could not recall the content of the previous questionnaires and interview (indeed, a small minority of subjects could not even remember being interviewed previously) due to the spacing of follow-ups at five-year intervals.

The re-interview of subjects also raises the issue of the first interview having a therapeutic effect on subjects. Again, any likelihood of such an effect is minimised by the five-year interval between interviews.

At base-line in 1978, subjects were only asked to nominate the number and duration of episodes in the past twelve months. There were no sex differences in these data. The rates for depressive cases with varying thresholds for case definition were detailed in Chapter 8, both for fixed periods and lifetime rates. There was a trend for both sexes to report fewer symptoms and fewer episodes over time, but there were no significant sex differences in rates for the higher thresholds (major depression +/- dysthymia). The second (lay) interviewer was slightly more likely to rate subjects as functioning well (reported in Chapter 7) but these differences were small. Higher rates by the first (clinician) rater may have been due to (i) the clinician being slightly more likely to rate dysfunction, (ii) subjects being found to be slightly less deviant at re-interview (Jorm, 1989), or (iii) some increase in confidence and maturation of personality by cohort subjects over the five years between interviews.

The findings based on data generated with the DIS on two occasions, five years apart, indicate that there are likely to be two trends leading to female preponderance. One of these is an artifact effect leading to a decrease over time in reporting rates for males, and the other is a possible increase in rates for females more evident over the latter five years (1983-1988), when there had been a wider diversity in social roles. So it is proposed that the difference is likely to reflect both a real phenomenon and an artifactual bias.

If the artifact hypothesis is supported, then how may such a bias operate? The sample have been noted as interested and motivated, in that members took part voluntarily and appeared pleased to be interviewed on each occasion - so that it is unlikely that the artifact can be attributed to sex differences in 'interview

characteristics' (whereby men might be more peremptory, impatient or dismissive at interview). The possibility that men might report fewer symptoms (for remembered episodes) is not suggested by the analyses in Chapter 9, with both sexes (and for females significantly) reducing symptom counts for episodes reported at each interview. It appears most likely that men may be more likely to forget actual episodes, which might be a direct sex difference, or a second-order effect emerging because the episode was less severe, less distressing, less disabling, less socially impairing or otherwise less 'memorable'.

There is no evidence in terms of duration, number of symptoms or help-seeking to suggest that the episodes occurring in males were less severe or distressing, but a recent paper discussing accessibility of sex difference in differentiation of psychiatric symptomatology (Briscoe et al, 1989) suggests such a possible mechanism. The authors note that males tend to be field-independent (i.e. analytical, perceptually discriminating, keeping experiences separate) while females tend to be field-dependent (i.e. do not perceive patterns as discrete entities). As noted previously, both subjects and witnesses tended to recall depressive episodes in terms of possible depressogenic life events and it may well mean that field-dependent individuals are more likely to recall life events and accompanying depressive episodes which cue recall in either direction while field-independent individuals separate the depressive episodes from life events, so that neither were likely to be recalled by prompting of the other. The other possibility is that subjects who "forgot" and were later diagnosed as 'new cases', failed to recall previous episodes because other episodes (either depression or in some subjects, panic disorder) had occurred subsequently.

While the numbers are small, the data may also indicate two different sub-groups of subjects reporting episodes of depression. First, those who became depressed at an early age (late teens or early twenties), who went on to have recurrent episodes, and another group who experienced a first episode of depression later (in late twenties or



thirties). The group with a younger age of onset would be similar to the young adults described by Weissman (Weissman et al, 1984), where high familial loading for depression (both genetic and environmentally induced) possibly plays a more dominant role than sex. The latter group may be more influenced by current life events (such as child-birth) and as such, one would expect sex differences to emerge in a cohort in their thirties leading to a female preponderance for either social or biological reasons (particularly associated with child-rearing) or a combination of the two.

### **Examination of sex differences**

#### **Socio-demographic variables**

There were increased numbers of married subjects at each wave, with a consistent increase in rate of parenthood, but importantly, the males and females continued to make these changes at the same rate. There was one predictable sex difference, in that some of the women left work during the early child-rearing years (particularly in the 1983-88 period). However, two men became chronically unemployed for psychiatric reasons (one with schizophrenia was excused from the study due to ill-health in 1988, the other experienced several episodes of depression and was involved in volunteer community work while receiving social service payments).

Findings of low levels of alcohol and drug consumption, and high levels of employment indicate that the effects of such potential confounding social risk factors such as chronic poverty, alcoholism and unemployment were minimised. A substantial group of females, but none of the men, left work to raise children which led to a sex difference in rates of full-time paid employment at ten-year follow-up. This difference in work patterns may have contributed to the trend towards female preponderance in rates of depressive episodes in the 1983-88 period.

### **Are the lack of sex differences due to a Type II error?**

There is sufficient power in the sample size to detect differences in the continuous variables where there is an effect size of greater than 0.45 standard deviations and to determine differences in the combined categories of depression if the sex differences were of the order expected (that is, a 2:1 female:male sex ratio). There is not sufficient power to reliably detect differences if the individual diagnostic categories are used (Cohen, 1977).

### **Were there sex differences in perceptions of social support networks?**

An interesting finding was of sex differences in perceived social support being greater during times of stress and that different help-seeking patterns evolved for each sex over time. Females consistently sought out family and friends in times of stress, over the decade of the study, while help-seeking patterns in males appeared determined by the most available source. Males increasingly turned to their partners for solace in times of stress (more so than females), while relinquishing the other avenues that females retained. Females reported higher levels of satisfaction with responses to requests for support in times of stress. There may be a relationship between satisfaction of perceived emotional support and recovery from depressive episodes, for females at least.

This finding supports a recent study examining gender issues, social support and recovery from depressive illness (determined by PSE) in 119 subjects (Brugha et al, 1990). They reported similar recovery rates from first onsets of depressive disorders but found that "living as married" and, to a lesser degree, number of social contacts (acquaintances rather than close friends) were positive predictors only for men, while negative interaction with their primary group (spouse, close family, good friends) was a negative predictor. For women, the base predictors of recovery were the number of primary group members named and satisfaction with social support. They raised the possibility that "men and women derive different benefits from their social networks, which may in turn relate to the future course of illness". The possibility of males using

networks of acquaintances, rather than close friends or family (Brugha et al, 1990) was not specifically examined in the teachers' study. Kessler and McLeod (1984) examined sex differences in rates of depressive symptoms (using CES-D and Zung's SDS) following negative life events. They found no sex differences in the effect of marital disruptions or events affecting spouses and children, but stated, that for men, "as soon as we moved beyond this small field of concern the impact of network events vanished, while it persisted among women". They reported that males showed greater distress than females after loss of income, and females, after ill-health, death of a loved one and network events (which involved persons other than spouse or children). They concluded that females were sought out more often in times of crisis than were males, knew more about events outside of their immediate family and considered interpersonal contact with their network as being more important than did males. These findings are concordant with the data from the present study.

The present data suggest that the females tended to maintain ties with friends and family, despite living with a partner and starting their own families, while the males became more reliant on their partners, particularly in times of stress. This is a finding that seems to be evolving over the years and, if confirmed, may explain the differential effects of marital state on the sexes. Overall, both males and females tend to go to females in times of stress (males, to their female partners; females to their female friends and family members). In young adult females, this is occurring at a time when they are often parents of young children themselves. These behaviours are in keeping with Gove's nurturant role hypothesis (Gove, 1984), that "the cost of caring" is increased depressive symptomology (Turner, 1988) and it does not seem surprising that the females reported increasing tiredness over the last five year period, when all of these factors were having an impact. However, these nurturant roles are not necessarily confined to women, and may increasingly involving men, particularly in middle-class, educated groups like this cohort.

### **How do the rates for DIS-derived depression categories compare with other studies?**

Studies reporting rates for depression categories using RDC and DSM-III criteria were discussed in Chapter 3. In comparison to large-scale general population studies, the numbers in this study were too small for detailed analysis of current prevalence rates and the discussion will focus on lifetime prevalence rates.

The lifetime rates in this study (by a mean age of 34 years) were 26%-32% for females, 17%-27% for males (for major depression alone); 31%-38% for females, 19%-32% for males (for DSM-III cases); 52%-60% for females, 36%-42% for males (for RDC cases), with the ranges of rates being predicated on which method of estimation of lifetime prevalence is used (as detailed in Chapter 9).

A New Haven study (Weissman & Myers, 1978) reported lifetime prevalence rates of 26% in females and 12% in males for RDC probable or definite major depression; 12% in females and 6% in males for RDC minor depression while 34% of the females and 17% of the males experienced one or both types of episodes. Reich et al (1980) determined a lifetime prevalence of major depression as 20-26% for females and 8-12% for men. In the ECA study, Robins et al (1984) calculated a 7% rate for females and 3% for males in lifetime major depression (being 5.5% for the whole sample and 8.7% in the sub-group aged 25-44 years). The rates for the individual centres are reported in Table 3.3. Additionally, they calculated a lifetime prevalence rate for dysthymia of 3.9% of the females and 2.0% of the males (the rate was 3.0% for the total group but 3.8% for those aged 15 to 44 years). The rates for major depression in the teachers' cohort are comparable to rates quoted in the first two studies but lower than those in the ECA study. The RDC case rates are higher in the teachers than in the New Haven study due to the contribution made by higher RDC minor depression rates. However, concerns about the low rates of minor depression in the New Haven study have already been raised in Chapter 3.

A study from Edmonton (Bland et al, 1988b) reported lifetime rates of 11.4% in females and 5.9% in males for major depression and 13.2% in females and 7.1% in males for DSM-III affective disorders. They also reported rates of 26.8% in females and 40.7% in males for "any core disorder", which includes anxiety and alcohol/drug abuse categories. They noted lifetime prevalence rates of 11.5% for major depression, 3.9% for dysthymia and 12.6% for affective disorders in the 25-34 age group (which is comparable in age to the teacher group) and the highest lifetime prevalence rates for phobias (35.4%) and substance abuse disorders (24.6%) and "any core disorder" (49.9%) in this age group. The rates for substance abuse are much higher than the teacher group. As the categories are not mutually exclusive, the rates for affective disorders and alcohol/substance abuse cannot be combined, but it is possible that for the 25 to 34 age group, the higher rates for depression in the teachers would be compensated for by the lower rates for alcohol and substance abuse. Rates for phobias are high in the Edmonton study. This category has been shown to be unreliable, as the DIS-derived data tend to over-report phobias (Myers et al, 1984; Burvill, 1987) or the diagnostic threshold for DSM-III criteria are set too low. The category for simple phobias was also the most unreliable anxiety disorder category in the teachers' study. The Edmonton group report a lifetime morbidity risk (Newman et al, 1988) of 22% in females and 16% in males for major depression alone, and 14.3% in females and 43.8% in males for alcohol abuse or dependence and 0.0% risk in females and 5.3% in males for antisocial personality. The question arises as to what these rates would be for both sexes in a group (such as the teacher cohort) with lower levels of alcohol abuse and sociopathy, and to what extent these high rates in males are masking affective disorders.

A Christchurch study (Wells et al, 1989) described rates of 16% in females and 9% in males for major depression only (see Table 3.3) and 19% in females and 10% in males for DSM-III cases. Their rates for anxiety disorders were obscured by combining agoraphobia and simple phobia and combining anxiety with somatoform disorders, which does not allow comparison with the teachers' data. For substance abuse disorder, they

report rates of 8.7% in females and 33.6% in males, and 68.5% in females and 63.0% in males for "any disorder covered". They also report lifetime prevalence rates for the 25 to 44 year age group, 15.5% for major depression alone, 5.5% for dysthymia, 16.8% for any affective disorder, 11.6% for anxiety/somatoform disorder, 22.6% for substance abuse disorders and 64.4% for any DIS-generated disorder. The rates for major depression and DSM-III cases in the teachers' study are higher but comparable while rates for dysthymia alone are lower. As in the Edmonton study, there are sex differences in lifetime rates for affective disorders but the presence of sex differences for "all psychiatric disorders" is determined by the inclusion or otherwise of alcohol or substance abuse and personality disorders.

Helgason (1986) summarised the position thus: "The comparison of expectancy rates between men and women shows that mental illness occurs with similar frequency among both sexes, but as different syndromes. This is reflected in the high expectancy of alcohol abuse and moderate expectancy of neuroses among men and high expectancy of neuroses and low expectancy of alcohol abuse among women. The obvious epidemiological hypothesis is that these disorders have some common etiologic factors....This hypothesis could be tested by studying a new cohort that has been subject to different attitudes towards alcohol and increasing per capita consumption of alcohol". The current cohort had low levels of alcohol consumption for both sexes, which may be reflected in higher rates of depression in the males.

Parker (1987) questioned the reliability of rates in the ECA study, but the points he raised are equally applicable to other studies using the DIS or similar case-finding instrument. He noted that the six-month prevalence rates were about half the lifetime prevalence rates and wondered whether the lifetime rates were accurate. He also questioned whether lay interviewers were as reliable as clinician interviewers and noted concern about the differences in rates between younger and older cohorts, postulating that older people may be less likely to endorse the questions about depression, for

cultural reasons (that they had not been encouraged to self-disclose), partly because they were less healthy and perhaps more likely to couch symptoms in terms of physical complaints.

Burvill (1987) had the same concern about the reliability of lifetime estimates, stating "personal experience and common-sense appraisal of the concept makes me sceptical that it is possible to estimate accurately the lifetime prevalence...when such estimates are made solely on the basis of a single structured interview with a respondent by a lay interviewer in a community survey, as was the case in the ECA Program". He was also critical that the authors of the ECA Program paper made no attempt to compare their rates to any other study, although it must be conceded that there were no other studies using the DIS at that time. He viewed the Lundby study (Essen-Moller et al, 1956) as the most comparable, and noted the advantages of a prospective study with a small team and access to supplementary information. He doubted the reliability of simple phobia and obsessive-compulsive disorder categories but concluded that affective disorder categories were likely to be among the more reliable. He compared the New Haven rates for four age groups for all DIS disorders (34.7% in 18-24 years old; 35.8% in 25-44 year olds; 22.2% in 45-64 year olds; 18.7% in those over 65 years) with the Lundby rates which reported a lifetime prevalence of psychiatric illness, which was 73.0% for females and 43.4% for males.

Since then, Bebbington's group (Bebbington et al, 1989) have also reported a very high morbid risk of depressive disorder (with entry criteria similar to minor depression), with rates of 70% for women and 49% for men to the age of 65 years. A longitudinal study of 998 community subjects (Amenson & Lewinsohn, 1981) used CES-D self-report measure, then SADS to determine depressive caseness. They concluded that "49% of the men and 62% of the women had experienced a diagnosable episode of depression some time during their lives". The last two studies report figures that are more comparable to the Lundby study and all use methodology requiring more than one estimation of

caseness and more extensive evaluation of the subjects in their personal context than the studies that have been quoted that used the DIS.

In 1985, Klerman stated that if he were to design a large community study again, he would include a screening self-report measure prior to using the case-finding interview with symptomatic subjects. Weissman (1988) has also questioned the use of the DIS as a stand-alone instrument and stated that "it is possible that full determination of lifetime rates requires more probing than is allowed in the DIS. Further evidence for this finding is that a large number of cases reported for a lifetime had occurred in the previous year...accurate determination of prevalence rates may require use of multiple informants and best estimates". Another group (McLeod et al, 1989) have advocated the use of a time-line to record depressive episodes and other key events (e.g. birthdays, holidays, crises) to aid recall. Spitzer's (1983) Longitudinal and All Available Data or LEAD standard had already been cited in Chapter 2 as a guide to the sort of information the researcher should consider gathering to make a reliable diagnosis of past episodes for psychiatric disorders.

While having one clinician interviewer (the author) recording data at both DIS interviews would have removed the problem of inter-rater reliability and differences between clinician and lay interviewers, it would have raised the problem of a 'halo effect' arising from prior knowledge influencing the manner in which the 1988 DIS interview was conducted. Even in a highly structured interview, it is possible to ask a question in a tone that suggests a positive or negative response. While some orienting information was provided (e.g. place and timing of previous interview, a particular hobby or interest of the person) to Robyn Curtain, the 1988 interviewer, intentionally no information was given that would determine whether the subject was likely to have been depressed. Inter-rater reliability was improved by the fact that both the author and Ms. Curtain were trained in the use of the DIS by Professor. Andrews' team, which had extensive experience with the instrument. Ms. Curtain was herself a trained nurse and psychology graduate, so that she



had more maturity and inter-personal experience than many of the ECA lay interviewers (who were frequently college students on holiday), and studies with volunteers and a random selection of study subjects indicated that inter-rater reliability was high (see Chapter 9). On both DIS assessment occasions, the instrument was given after a semi-structured interview that covered a wide range of topics from employment and health issues to social support networks and coping styles. At both interviews, a time-line was used to record all episodes of depression and any life events offered, which may have enhanced memory recall and led to a more accurate estimation of lifetime prevalence of major and minor depressive episodes.

The use of two interviewers over a relatively long period allowed for three different methods of estimation of lifetime prevalence (reported in Chapter 8). It is assumed that the real rates are somewhere within the range quoted and summarised early in this chapter. All three methods give rates that are comparable (for major depression) to the New Haven (Weissman & Myers, 1978) study and, in females, also to the Christchurch study (Wells et al, 1989). When minor depression is included, the lifetime rates are similar to those for the Lundby (Essen-Moller et al, 1956), Midtown Manhattan (Srole et al, 1978) and Camberwell (Bebbington et al, 1989) studies. Indeed, there is considerable consensus that about 40% of the population will be categorised as 'well' with the remaining 60% being categorised as mildly to severely impaired (in Midtown Manhattan study) or at least mild depression (Lundby study) or borderline episodes of depression (Camberwell study).

In the teachers' study, diagnoses were moderately stable over time. Inconsistencies from one assessment to the other were largely due to (i) changes in reported numbers of symptoms (so that the depressive experience was recalled but there was a change in diagnostic category) or (ii) changes in attribution (so that the incident was recalled on both occasions but only noted as "having a significant impact on life" or as constituting depression on one assessment). Dysthymia was relatively uncommon and

did not exceed rates from other studies, but this category was the most inconsistently recalled depressive category. The episodes generally lasted just over two years and were often termed minor depression if reported five years later, when reported episodes later failed to reach the minimum two years' duration. Often professional help was not sought for such episodes. There has been some concern over the heterogeneity of this category (Akiskal, 1983; Kocsis & Frances, 1986) and its high co-morbidity with other medical and psychiatric conditions (Weissman et al, 1988). Those two groups have noted that the category may include unresolved major depression and long-lasting characterological and non-melancholic depressions, otherwise subsumed under the rubric of 'neurotic depression'. In the teachers' cohort, those subjects who attracted the diagnosis usually reported episodes in their early 20's following a relationship break-up, at a time when they were unsure of re-evaluation of their life goals. These episodes fulfilled the criteria but probably not the spirit of the diagnosis of dysthymia (which implies minor depression for much of the time, for a minimum of two years) but were more in keeping with the diagnosis of RDC intermittent minor depression which implies episodes of minor depression fluctuating over at least a two-year period. The difference in these concepts and earlier comments on heterogeneity all reflect the lack of clarity in defining this category.

### **Why are lifetime prevalence rates higher in younger cohorts in recent times?**

Parker's (1987) questioning of the higher rates of depression in the younger-age cohort in the ECA study also coincided with reports of a cohort effect of increasing rates of depression, and decrease in sex differences in rates in younger groups (Srole & Fischer, 1980; Hagnell, 1982; Klerman et al, 1985; Murphy, 1986). All these latter groups judge there to be a real change in rates, with young males experiencing increased rates of depression, with a possible decline in expression of anxiety symptoms, while rates for depression in females have either remained stable or declined. It is also possible that this is an artifactual change, in that males may be becoming more willing to discuss their symptoms, or may be seeking help rather than 'self-medicating' with alcohol. It is

too early to know which is the correct explanation but the findings in the male teachers reflect this phenomenon of increased rates in males.

Another finding in two longitudinal studies (Srole & Fischer, 1980; Murphy, 1986) was that, while rates of depression for females had risen quickly during the 20-35 age range, females in their forties reported declining rates of depression not evident in males. However in the Midtown study, there was a methodological problem in that some of the women who had married and changed their names were the most difficult to trace, so that the sample may not have been representative (Srole, 1980).

It is likely that there are a number of explanations for the finding of increased rates of depression in younger adults. Firstly, that depression is not sufficiently recognised in the elderly, partly due to problems already mentioned, including (i) inaccurate recall of lifetime prevalence, (ii) the possibility that older people being more likely to attribute their distress to physical rather than psychological causes (when the DIS probes would exclude such symptoms), or (iii) their being less accustomed to self-disclosure of symptoms (Hasin & Link, 1988), or (iv) that there is a bias in the wording of the constructs in the DSM-III criteria towards depressive experience in young people.

Secondly, the criteria for DSM-III affective disorders have shorter minimum duration for depressive disorders than for anxiety disorders, giving the potential for episodes with mixed features of anxiety and depression to be more readily categorised as depressive disorders. Kendell (1989) has pointed out that most clinicians assume that depressive disorders take precedence over anxiety disorders in such hierarchies and a hierarchical approach is implicit in DSM-III. He stated that "our habits of thought are now so strongly based on hierarchical assumptions that we are often unaware of the full range of the patients' symptoms". For this reason, it is important to use standardised diagnostic interviews such as the DIS and SADS, and to compare rates derived from these interviews with those from other sources, such as clinician interviews and family

history sources. The same diagnostic hierarchy can be applied or disregarded when DIS/DSM-III algorithms are used. When exclusion criteria were ignored in an examination of the ECA data (Boyd et al, 1984), there was "a general tendency towards co-occurrence, so that the presence of any disorder increased the odds of having almost any other disorder, even if DSM-III does not list it as a related disorder". Thus the rates for depressive disorders may, at least to some extent, be artifactually raised by use of the DSM-III system, although this does not explain changes in rates in the Lundby study (Hagnell et al, 1982). The overlap between anxiety and depressive disorders will be discussed later in this chapter.

If the findings of increased rates in younger cohorts are not artifactual, possible explanations include secular changes to, or increased exposure in current depressogenic stressors, increased rates of help-seeking in the young (reinforcing recall), or more individuals with a high risk for depression becoming parents themselves (e.g. the advent of such effective treatments as tricyclic antidepressants may have allowed more people genetically predisposed to depression to produce offspring who were in turn predisposed to depression).

### **Have there been differences in social role expectations for younger cohorts?**

If the equal rates of help-seeking for the male and female teachers in any way reflect the wider population of young adults, then it is possible that males are seeking more help for their emotional problems, which could be reflected as increased rates of depressive disorders in males, either because of an actual increase in rates of depression or increased awareness and recall of such episodes.

Srole and Fisher (1980) interpreted the findings on changing rates between the sexes in terms of increased mental health in "a new breed of women" and proposed that even from one generation to the next, society can have very different expectations of both sexes. They concentrated on womens' increasing penetration into previously male-

dominated professions and occupations, and suggested that this new-found freedom was associated with declining rates of anxiety and depression for women. This finding is in keeping with other studies which have shown that women involved in employment away from the home have less experience of depression (Weissman et al, 1973; Radloff, 1975; Brown & Harris, 1978; Cochrane & Stopes-Roe, 1981). The changing status of women could be expected to have an impact on men. Kessler and McRae (1982) found higher rates of depressive symptoms in men whose wives worked outside the home compared to those whose wives were solely engaged in home duties, but for those with working wives, the husbands who helped their wives with the care of the children experienced fewer symptoms than those who left these tasks solely to their wives. These findings are affected by the very issues of care and control that the IBM was designed to quantify, but draw attention to the changes for men, from sole breadwinner to a role of 'partner' in a symmetrical relationship. These roles are still evolving in contemporary society (Young & Willmott, 1973; Murphy, 1986) and this cohort is part of the generation involved with such issues. The finding that mean scores on the femininity and masculinity subscales of the sex role inventory (Bem, 1974) reflected the concept of androgeneity for both sexes is in keeping with the hypothesised social role changes and the finding that both sexes endorsed similar reasons (whether primarily masculine or feminine in sex-role typing) for teaching as a career choice.

So the largely negative findings in terms of significant sex difference in lifetime and extended period prevalence data were against initial expectation but consistent with the cohort's professional utilization rates (assessed both at base-line and reviews in 1983 and 1988), supported by corroborative witness data in 1983 and in keeping with the trends reported from the other longitudinal studies quoted. Equally important, there were virtually identical incidence rates for the males and females for depressive disorders over the first five years of the study when both sexes were involved in social role transition but a non-significant trend for higher rates of depression in females from 1983-88,

measured at the 10-year follow-up, which coincides with females raising children and leaving work (i.e. greater social role disparity between the sexes).

Weissman and colleagues (1984) reported a median age of onset of 27 years for major depression, which was the mean age of the cohort at the 1983 assessment when rates of depression were high for both sexes. However, Jorm (1987) showed a sex difference in rates of depression emerging in adolescence and being pronounced in the twenties but has suggested that the sex difference does not peak until the early thirties and the peak may occur later in a middle-class group (e.g. by delaying birth of first child). Thus a sex difference may be starting to emerge, either due to the effects of rearing children (whether biological or simply of being at home with young children) or effects of marriage and patterns of social support.

The other possibility is that, as social variables effectively remained controlled throughout the study (in that age, marital status, parenthood, and employment apart from home duties) did not differentiate the sexes, a sex difference was prevented from emerging.

### **What is the relationship between anxiety and depression?**

Studies reporting the increasing rates of depressive disorder in young people, particularly males, and possible declining rates of anxiety disorders (Hagnell et al, 1982; Klerman et al, 1985; Murphy, 1986) have been noted earlier. The question of whether this finding is partly an artifact (due to changes in emphasis in psychiatric diagnosis) was raised (Murphy, 1986). Eaton and Ritter (1988) stated that the DIS-derived data for anxiety and depression from the ECA study were analysed in parallel fashion, but when the data were analysed together, only slight separation between the two sets of syndromes was evident. They questioned whether this lack of separation was due in some way to use of the DIS. Earlier, the Bedford College group (Finlay-Jones et al, 1980) had found cases of anxiety (free-floating or situational anxiety or panic attacks)

and depression (depressed mood and a number of depressive symptoms) generated from PSE data in a community sample to be mutually exclusive.

A likely explanation for this apparent discrepancy was put by Winokur (1985) when he noted that European psychiatrists have always categorised neurotic depression on a phenomenological basis, whereas in the 1940s to 1970s, American psychiatrists had turned to a dynamic-interpretive basis, which was found to lack diagnostic validity. He stated that DSM-III major depression was a lowest common denominator seen under multiple syndromes (e.g. primary depression, alcoholism, schizophrenia) and that "DSM-III essentially regards all depressions as equal and attempts to separate them with a separate axis (i.e. personality disorders)." The same might equally be said for the anxiety disorders, so that it may well be the use of the DSM-III criteria, rather than the DIS, which is causing problems in differentiating anxiety and depression for American researchers, while European researchers, generally using ICD-9/PSE criteria do not seem to have encountered the same difficulties.

A review of studies of lifetime occurrence of anxiety and depression (Brier et al, 1985) suggested that most researchers identified panic disorder and agoraphobia as discrete categories, and reported that they often coexisted with major depression, but found less evidence to support a separate category of generalised anxiety disorder. Leckman et al (1983) found an increased risk of major depression, panic disorder and generalised anxiety disorder in relatives of those with major depression and panic disorder over those with major depression alone and postulated a shared, familial diathesis to major depression and panic disorder. Reich (1986) summarised his review of the literature thus: "A reasonable middle position to take would be that in many cases anxiety and depressive disorders can be distinguished by symptom complex and course of illness...those patients who are suffering from both anxiety and depressive disorders are at especially high risk for both disorders and a poor prognosis." The last point had previously been made by Murphy (1985) in her reanalysis of the Stirling County data.

In this study, anxiety disorders generally were co-occurrent with or post-dated the onset of depression. There was no instance of panic disorder (with or without agoraphobia), or generalised anxiety disorder in subjects who had not experienced at least an episode of minor depression. These findings fit with the reported overlap between anxiety and depressive disorders but have meant that the category of RDC cases (when all RDC depressive categories were combined) has also incorporated all of the major anxiety disorder categories, as well as depressive categories and some who may have been classified as 'neurotic' or 'anxiety disorder' if another (unspecified) set of diagnostic criteria were in operation. Whatever the name, the 'RDC cases' grouping seems to have encompassed those cases with a general neurotic diathesis, such as the "general neurotic syndrome" proposed by Andrews' group (1990), which is an interesting finding in itself. This finding also implies a primary mood disorder in neurotic depression, rather than seeing all such episodes as maladaptive responses to anxiety (Wolpe, 1986).

### **What is the relationship of life events to onset of depression?**

This is another area where there has been a large body of research which will only be dealt with briefly. As the life events scale of Holmes and Rahe (1967) simply rank-ordered potentially distressing life events, researchers moved to examining more subjective ratings of the impact of life events on the individual, and the effects of specific life events, such as child-birth, nuclear disaster (Bromet et al, 1986), unemployment (Jacobson, 1987; Bolton & Oatley, 1987). Such research established that the effect of the life event will be influenced by other determinants both practical (e.g. Jacobson found that the impact of unemployment was determined by available finance), and psychosocial (e.g. Bolton and Oatley found that depression was likely in unemployed men, if job loss also constituted loss of social support that was not replaced).

The Bedford College team have made continued contributions in the area of life event research. The group has come to rely on interviewer-generated assessment of life



events (which are then appraised by the group to give an objective rating of impact on the subject's life) rather than self-report measures (Brown et al, 1987). Additionally, they have postulated that 'danger' events precipitate anxiety and 'loss' events precipitate depression (Finlay-Jones and Brown, 1981). It would be very important to be able to separate anxiety and depressive disorders if these concepts were to have any validity.

Brown's group (Brown et al, 1986) later added the concept of being 'let down' by important figures in the subject's social support network, which they hypothesised as a risk factor for depression but noted that an inherent difficulty in assessing the importance of the construct is that the person involved in the 'letting down' was often the same person who was the cause of the distress. They have also examined two social factors, reduction of a severe difficulty (e.g. a chronically unemployed husband starts work) and a 'fresh start' event (e.g. a separated woman gains a divorce and moves into a comfortable home) which are both reported to be statistically significant in relation to recovery from depression (Brown et al, 1988).

Thus research effort has now moved to determining the availability of social support in times of stress, with confiding relationships being considered to be the most protective (Surtees, 1980). Some suggested gender differences in help-seeking patterns have been discussed earlier (Brugha et al, 1990).

In the teachers' study, life events were controlled to some extent as, in 1979, the entire cohort left Teachers' College to start work and both sexes took up other adult roles of marriage and parenthood at similar rates. Those who were cases in 1983 (using all three diagnostic groupings) had experienced more negative life events in the twelve months preceding the 1983 assessment. There was no evidence of plaintive set nor prior sensitivity to life events in cases, as the estimated impact scores were similar to the actual weightings given when events occurred both for cases and non-cases. Life events were

seen as contributing to the timing of depressive episodes rather than as vulnerability factors, which is in keeping with the Bedford College model.

**How do the data from the study compare with the Bedford College model for social factors related to depression?**

The work of Brown, Harris and their colleagues has been already noted in Chapters 1 and 10. Harris (1989) has recently reviewed the scope of the group's work over the past 15 years, from which has evolved a comprehensive model for social origins of depression (which also acknowledges the importance of psychological and biological factors). The group generated definitions of caseness, based on the PSE, with 'definite' and 'borderline' case categories. They conceptualised 'cases' as subjects for whom clinical intervention would be appropriate, at least at the level of psychiatric outpatients, and 'borderline cases' as those subjects manifesting social difficulties where intervention would not necessarily be required (Brown et al, 1985). While the PSE was used as the case-finding instrument, the diagnostic criteria have been defined so as to be comparable with DSM-III and RDC categories, although there are no minimum duration criteria for episodes (Finlay-Jones et al, 1980).

They then examined factors leading to the onset of depression, identifying four 'vulnerability factors' which seemed to potentiate the role of 'provoking agents', namely absence of a husband or boy-friend who was also a confidant, loss of mother prior to age of 11, presence of 3 or more children in the home of 14 years or less, and lack of paid employment outside the home. Since then, the first factor has been found to be the most robust and also coincides with Gove's findings that the quality rather than the presence of a relationship with a husband or live-in partner is the important determinant of depression (Gove et al, 1983). The effect of the second factor was later shown to be influenced more by the quality of parental care and circumstances before and following the loss (Parker & Hadzi-Pavlovic, 1984; Harris & Brown, 1985). The effect of the other two factors (lack of employment and presence of 3 or more children at home) may also

depend on the context and may reflect underlying personality, as well as social factors. Brown and Bifulco (1990) hypothesised that low self-esteem, or 'negative evaluation of self', may be an important underlying determinant but have not attempted to quantify the construct. They link both early separation from a parent and lack of close confidant with poor self-esteem, suggesting a model whereby poor unsupportive early relationships lead to long-term cognitive vulnerability to depression (low self-esteem) and further vulnerability in the greater likelihood of having unsupportive relationships in adult life. In the presence of a provoking agent, these factors lead to lack of hope in self or others, with subsequent generalisation of hopelessness and depression. The possible restorative 'reduction of difficulty' or 'fresh start' life events have already been noted. Harris concluded that there may be gender differences in that women may develop their sense of identity from emotional rather than functional relationships, which again echoes Gove's (1984) differentiation of nurturant and fixed roles.

A group in Edinburgh have tested some social vulnerability factors in a community sample of 376 women (Surtees et al, 1983, 1986). They used the PSE with additional questions from SADS, so that CATEGO and RDC diagnostic criteria could be generated to define caseness. They noted 7 vulnerability factors, lack of close social support, lack of diffuse social support, poor relations with household members, living alone, an unemployed spouse, loss of a parent prior to 11 years of age, previous professional help-seeking for 'nerves' and low self-esteem (using the Rosenberg self-esteem measure which was the instrument in the present study). They reported that prior episodes of depression were a more significant predictor of subsequent episodes of depression than low self-esteem (Ingham et al, 1986) and later reported that major depression was best predicted by "an interaction between total stress and low self-esteem" (Miller et al, 1989). Henderson's group (Henderson et al, 1981) in Canberra have also examined social support in cases of depression and anxiety (derived by PSE/CATEGO). The findings concerning the importance of perceived availability of support have become influential in terms of discriminating between support in general

and specifically in times of stress, and between actual and perceived support. Henderson later stated (Henderson & Brown, 1989) that the study should have included more measures of personality, rather than relying solely on the EPI neuroticism scale. The Edinburgh and Canberra groups have made significant contributions to the body of knowledge in the social psychiatry field but have not come up with such a comprehensive model of depression as the Bedford College group.

The definite depressive case definition used by Brown's group is similar to definition of DSM-III major depression and is roughly comparable to rates for the DSM-III cases in the teachers' study, while RDC cases are roughly comparable to the combined categories of definite and borderline cases for anxiety and depression, so that comparisons between the two studies are possible. Indeed, the one-year prevalence rates for 363 Camberwell women (Brown et al, 1985) were 17.6% for definite cases of depression and 49.0% for definite and borderline cases of anxiety and depression (excluding simple phobias), so that their twelve-month prevalence rates show a similar pattern to lifetime prevalence rates (to the age of mid-30's) for the teachers. The teachers in this study are at the other end of the social spectrum from the working class women of Camberwell but the studies are similar in that both groups were selected for their social homogeneity. As men were excluded from the Camberwell investigation, one can only postulate on the effects of the proposed vulnerability factors on them. It is highly likely that being unemployed and having three children at home would also be risk factors for Camberwell males and that as in the teachers, low self-esteem would be a risk factor for both sexes.

The present study has not concentrated much on identifying current social vulnerability factors but more on those risk factors that were evident, as much as possible, prior to the onset of the depressive episodes. For the teachers, the consistent findings concerning perceptions of childhood experience, with low parental care (particularly paternal) and high overprotection (particularly maternal) are postulated as

precursors to low self-esteem and to dysfunctional intimate relationships in adult life. One could postulate that those subjects whom Brown reported as suffering adversely from effects of parental death or separation may well have also returned scores indicating at least low parental care (with high overprotection being more dependent on individual circumstances) if the PBI had been used. Low self-esteem is seen as a significant risk factor with an hypothesised relationship between self-esteem and inadequate perceived parental care. The fact that these variables are found to be significant in two very different social groups using different methodology (but somewhat similar definitions of caseness) increases the possibility that these variables have an important part to play in predicting vulnerability to depression.

The findings reflect Bowlby's (1973) attention to the importance of secure attachment to an important figure in childhood to provide a basis for self-reliance as an enduring trait. Robson (1988) has examined the concept of self-esteem and found that it was an idea that was conceptualised in a variety of ways. He concluded that there was a definite association demonstrated between self-esteem and clinical disorder but that research had been hampered by lack of clearly conceptualised scales to measure self-esteem, and noted a need for prospective studies to disengage the effect of depression on self-esteem. Robson's criticism of Rosenberg's (1965) scale has already been noted in the previous chapter. The improvement in self-esteem (as measured by the Rosenberg scale), after the cohort started work could reflect such constructs as personal worth and social competence. If self-esteem is an internalisation of the parental environment, the PBI itself may capture these constructs better.

The effects of high trait neuroticism may be independent but additive, particularly in times of stress. Brown's group do not have measures comparable to the trait depression, dependency and sex-role scales, so that no comparisons are possible.

### **What is an appropriate definition of caseness?**

The importance of definitions of caseness for identification of subjects with depressive disorders rather than distress was discussed in Chapter 1. Besides the obvious research and clinical considerations, there are also political and feminist considerations that have been raised in connection with the derivation of DSM-III classification system. Russell (1986) has stated that the focus of DSM-III is on the problems of individuals rather than the social context, so that treatment of the individual is encouraged, rather than social change. This is a valid concern that research in the area of social psychiatry tries to address while underlining the importance of using caseness definitions appropriate to the question at hand. The feminist concerns (also touched on in Chapters 1 and 4) include the differing yardsticks for mental health in males and females, and need for attention to wording of questions, possibility of different rates of recall of episodes, and numbers and types of symptoms generated by either sex. All of these factors will have an effect on prevalence rates and generally tend to increase rates in females.

The current study used the DIS as a case-finding instrument from which three definitions of caseness have been considered, two of which are based entirely on data derived from the DIS-generated categories. The combined RDC categories, termed RDC cases (major depression, definite episodic minor depression and definite intermittent minor depression), and DSM-III cases (significant major depression or dysthymia) provided categories that were comparable to other studies. The DSM-III case categories could be compared with the three studies of general population groups where data were also available for sub-groups of similar age. The RDC case grouping provided a lower entry threshold which may have been relevant in an examination of sex differences. If women complained of more symptoms at a sub-clinical level, one would expect more females to be included as the threshold for caseness was lowered. The RDC impairment criteria incorporated in the DIS also made some provision for defining caseness in terms of help-seeking behaviour, which is also important in investigating sex differences in rates of depression in a community sample. These issues of help-seeking were pursued

independently in more detail in the study and it is suggested that the RDC impairment criteria require further operationalisation, particularly in relation to the question as to whether the episode has had a 'significant impact' on the subject's life. The attention to social role impairment applied by Angst's group (1984c) and Keller's group (1987) are appropriate to general population groups, and provide models for defining impairment that could be incorporated in future studies.

A third 'defined cases' method was based on the DIS-generated categories, with additional criteria that amplified the RDC functional impairment appropriate to this cohort. The criteria used for 'defined cases' were listed in Chapter 8 and were designed to delineate the group whose depressive episodes had caused significant disruption to their normal role function and that may have required psychiatric intervention. This method allowed for the inclusion of subjects who had not sought help, if they had taken a prolonged period away from work, attempted suicide or had a sustained episode of depression (reported on at least the first assessment occasion with a minimum duration of twelve weeks with at least two symptoms). In this way, there was an attempt to operationalise psychosocial impairment in a fashion that was relevant to cohort members of both sexes and compatible with the information elicited. The RDC case rates were the highest, the DSM-III rates the lowest, and the 'defined case' rates fell between the two.

None of these case-finding methods was intended to derive cases of melancholic depression, as it was expected that this depressive diagnosis would be relatively uncommon in such a cohort.

Vaillant and Schnurr (1987) had found that the presence of a DSM-III diagnosis was a "valid categorical definition of a case", while Kendell (1989) has commented that "validation depends on the elucidation of aetiological processes". For the teachers, a number of predictor variables were able to discriminate consistently between cases and non-cases defined with DSM-III and RDC criteria, although extra criteria (for 'defined

cases') were shown to enhance the definitions of caseness. Thus the same risk factors were of relevance whichever method of case definition was used, although the third method ('defined cases') led to the highest rates of correct classification and greatest discrimination between cases and non-cases. The 'defined' caseness method demonstrates a useful combination of the DIS-generated data with further criteria applied that were relevant to this group, and was successful in increasing the correct classification rates for cases and non-cases in the discriminant function analyses.

### **How important was gender as a risk factor to depression for the caseness definitions used in this study?**

Gender was irrelevant as a risk factor for DSM-III and 'defined' cases, but had some relevance for RDC cases (where the diagnostic threshold was lowest).

The lack of significance of sex as a risk factor shows the importance of comparing the sexes for evaluation of risk factors. Sex biases should also be taken into account when selecting definitions of caseness and consideration of length of time period under investigation.

### **Conclusions**

The findings from the study vindicate the original aim of studying a homogeneous group in that a number of potentially confounding risk factors (such as economic privation, substantial alcohol and substance abuse, unemployment, gross sex differences in expectations from careers and relationship opportunities) were minimal.

There were no sex differences in rates of depressive experience at base-line, allowing for acceptance of the first hypothesis.



There was satisfactory consistency and reliability of data gathered at five-year intervals. Differences in information recall by subjects was an important factor in differences, which were partly sex-related.

Rates for depressive disorders were high and equal for both sexes. When compared to other studies, rates were comparable, particularly for females, while male rates may have reflected the relatively low rates of alcohol abuse and sociopathic personality disorder in this group of males, and the possibility that teaching as a profession is favoured by males who are more nurturant when compared to other general population groups in the same age range.

Trends towards sex differences in rates increased as the criteria for depressive caseness were lowered. There were sex differences in lifetime prevalence rates for RDC cases (the category with the lowest entry criteria). Lifetime prevalence rates were also influenced by the length of time from experience of episode to interview. Thus, the second hypothesis was largely disallowed.

The generally negative findings do establish that it is unlikely that biological factors have any direct effect on determining sex differences in mild to moderate depression. The theoretical relevance of possible genetic or hormonal factors to any sex difference has been considered in Chapter 4. Any biological determinants are likely to have greatest impact around child-birth. Here there has been a modest rise in rates for the 1983-88 period, when more females were having a second child and leaving full-time employment, but the overall rise in new case rates fails to reach statistical significance (although this could be due to low rates). Overall, this does not constitute a compelling argument for biological causation but biological factors may still have an effect.

The longitudinal design means that data were collected from subjects, in most cases, prior to their first depressive episode. However, there was a group of subjects who

had been depressed prior to the commencement of the study and data were analysed both including and excluding this subset. Despite losing some of the benefits of the longitudinal design, it was deemed important to include them in many of the analyses as subjects in this subset were among the more severely affected by depression. Where analyses included only 'new cases after 1978', the results were virtually unchanged, however, some of the more depressed subjects were excluded which lessened the ability of the variables to discriminate between cases and non-cases.

The risk factors for mild to moderate depression include negative perceptions of early parental environment (especially low paternal care and high maternal over-protection), low self-esteem, high dependency and high trait depression. Negative life events are conceptualised as precipitating events, which must then be viewed in the context of availability of adequate social support, particularly in times of stress, which may be different for each sex. Absence of a confidant constituted a risk, as did the presence of an unsupportive intimate relationship characterised by low care and/or high control, the former being of more importance in females and the latter in males. In this group, the presence of three or more young children at home constituted an added risk for both sexes.

The risk factors hypothesised as being important for female vulnerability to depression (Hypothesis 3), such as high interpersonal dependency, high femininity scores and presence of three or more children at home were found to be risk factors for both sexes, while there was some support for exposure to motherhood and home duties in females. Hypothesis 3 was partially supported in that the risk factors were shown to be relevant to both sexes, and more so for males rather than solely for females.

Risk factors considered to be associated with depression for both sexes (low perceived care and high overprotection from parents, high neuroticism, low self-esteem, low care and high control from partner) were found to be risk factors for depression.

Hypothesis 4 was accepted.

### Postscript

A ten-year study affords time for reflection and change in the research context in which the study is placed. Over this time, the basic concepts of depression are still being examined and undergoing change (Parker et al, 1990) and a further edition of ICD, namely ICD-10 (W.H.O. 1990) is now in draft form and a further edition of DSM, namely DSM-IV is in planning stages. There has also been opportunity for researchers to take stock of the utility of structured case-finding instruments and they have been found to be reliable, particularly if the data can be checked by multiple assessments or other sources. The findings concerning reliability of data over long periods question the reliability and validity of lifetime prevalence data gathered at one interview only.

Large-scale epidemiological studies can provide useful data on prevalence rates, while studies of smaller numbers provide data which has more depth and takes note of subjective experience. Both have their place and the use of similar case-finding techniques and instruments across studies allows for a pooling of research findings.

Longitudinal studies need to commence early in life if new cases are being considered and, as a corollary, some of the most severely affected subjects may be those that present earliest. The most severely affected are often more difficult to contact and may also have more disturbed relationships with their family of origin and be less likely to have a partner or a well-functioning intimate relationship, and these issues require consideration in research design.

The high rates of depression when caseness threshold is lowered to include minor depression or borderline cases (with lifetime prevalence rates of about 60% from a variety of studies) call into question the meaning of these concepts and the relationship between anxiety and depression.

As a psychiatrist, it has been interesting to follow a group of normal young

people, to examine how such a group copes with life and how those with serious episodes of depression cope with or without professional intervention. There is no doubt that examining both sexes simultaneously added to the depth of the study. It is my intention to continue following the group and I would predict that those likely to have depressive episodes in the future have generally 'declared themselves' by now, at least until the cohort reach the 50-60 year age group, when a small number may report episodes of melancholic depression. Theoretically such episodes would be more likely to be associated with a positive family history of melancholia or bipolar disorder, older age and physical illness and could occur equally in cases or non-cases but that is an interesting research question in itself.

Other questions that can be examined with a longitudinal design include (i) the possible evolution of sex differences in social support and marital styles, with interaction between gender, marital state and depressive caseness, (ii) the relationship between perceived parental style and evolving relationships with partner, particularly whether marital relationships come to resemble perceived parental relationships over time, (iii) the long-term consistency of self-report measures, (iv) the effects of the presence of small children and changes in work patterns on parents of both sexes, and (v) in this group, there is the possibility of using the data to improve the lot of the teaching profession in Australia.

## REFERENCES

- Akiskal, H.S. (1983). Dysthymic Disorder: Psychopathology of proposed Chronic Depressive Subtypes. American Journal of Psychiatry, 140, 11-20.
- Akiskal, H.S., Bitar, A.H., Puzantian, V.R., Rosenthal, T.L., Walker, P.W. (1978). The nosological status of neurotic depression. Archives of General Psychiatry, 35, 756-766.
- Alloway, R., Bebbington, P. (1987). The buffer theory of social support - a review of the literature. Psychological Medicine, 17, 91-108.
- American Psychiatric Association (1980). Diagnostic and Statistical Manual of Mental Disorders, Third Edition. Washington, DC:APA.
- American Psychiatric Association (1988). Diagnostic and Statistical Manual of Mental Disorders, Third Edition, revised. Washington, DC:APA.
- Amsenson, C.S., Lewinsohn, P.M., (1981). An investigation into the observed sex difference in prevalence of unipolar depression. Journal of Abnormal Psychology, 90, 1-13.
- Andreasen, N.C., Endicott, J., Spitzer, R.L. et al (1977). The family history method using diagnostic criteria: Reliability and validity. Archives of General Psychiatry, 34, 1229-1235.
- Andreasen, N.C., Grove, W.M., Shapiro, R.W., Keller, M.B., Hirschfeld, R.M.A., McDonald-Scott, P. (1981) Reliability of Lifetime Diagnosis, A Multicenter Collaborative Perspective. Archives of General Psychiatry, 38, 400-405.
- Andreasen, N.C., Winokur, G. (1979). Newer experimental methods for classifying depression. Archives of General Psychiatry, 36, 447-452.
- Andrews, G., Neilson, M., Hunt, C., Stewart, G., Kiloh, L.G. (1990). Evidence for a general neurotic syndrome. British Journal of Psychiatry, 157, 6-12.
- Angst, J., Dobler-Mikola, A., Binder, J. (1984a). The Zurich study - A prospective Epidemiological study of depressive, neurotic and psychosomatic syndromes. 1 Problem Methodology. European Archives of Psychiatry and Neurological Sciences, 234, 13-20.

- Angst, J., Dobler-Mikola, A. (1984b). The Zurich Study. 3. Diagnosis of Depression. European Archives of Psychiatry and Neurological Sciences, 234, 30-37.
- Angst, J., Dobler-Mikola, A. (1984c). The definition of depression. Journal of Psychiatric Research, 18, 401-406.
- Anthony, J.C., Folstein, M., Romanoski, A.J., Von Korff, M.R., Nestadt, G.R., Chahal, R., Merchant, A., Brown, C.H., Shapiro, S., Kramer, M., Gruenberg, E.M. (1984). Comparison of the Diagnostic Interview Schedule and a standardized psychiatric diagnosis: experience in Eastern Baltimore. Archives of General Psychiatry, 42, 667-675.
- Antill, J.K., Cunningham, J.D. (1979). Self-esteem as a function of masculinity in both sexes. Journal of Consulting and Clinical Psychology, 47, 783-785.
- Arrindell, W.A., Bodens, W. and Lambert, H. (1983a) On the psychometric properties of the Maudsley Marital Questionnaire (MMQ): Evaluation of self-ratings in distressed and 'normal' volunteer couples based on the Dutch version. Personality and Individual Differences, 4, 293-306.
- Arrindell, W.A., Emmelkamp, P.M.G. and Bast, S. (1983b) The Maudsley Marital Questionnaire (MMQ): A further step towards its validation. Personality and Individual Differences, 4, 457-464.
- Arrindell, W.A., Schaap, C. (1985) The Maudsley Marital Questionnaire (MMQ): An extension of its construct validity. British Journal of Psychiatry, 147, 295-299.
- Atkinson, A.K., & Rickel, A.U. (1984). Post-partum depression in primiparous patients. Journal of Abnormal Psychology, 93, 115-119.
- Baldwin, R.C., Jolley, D.J. (1986). The prognosis of Depression in Old Age. British Journal of Psychiatry, 149, 574-583.
- Barnett, P.A., Gotlib, I.H. (1988). Psychosocial Functioning and Depression: Distinguishing among Antecedents, Concomitants and Consequences. Psychological Bulletin, 104, 97-126.
- Barrett, J. (1986). Case Identification for Category Validation: The Challenge of Disorder-Specific Assessment. Comprehensive Psychiatry, 27, 81-100.
- Barrett, J.E., Barrett, J.A., Oxman, T.E., Gerber, P.D. (1988). The Prevalence of Psychiatric Disorders in a Primary Care Practice. Archives of General Psychiatry, 45, 1100-1106.

- Bebbington, P.E., (1988). The Social epidemiology of clinical depression. In Handbook of Social Psychiatry, eds. Henderson, S., Burrows G. Canberra: Elsevier.
- Bebbington, P. (1987). Marital Status and Depression: A Study of English National Admission Statistics. Acta Psychiatrica Scandinavia, 75, 640-650.
- Bebbington, P. (1989). Social epidemiology of depression and the neuroses. Current Opinion in Psychiatry. 2. (204-207).
- Bebbington, P., Hurry, J., Tennant, C., Sturt, E., Wing, J.K. (1981). Epidemiology of mental disorders in Camberwell. Psychological Medicine, 11, 561-579.
- Bebbington, P., Katz, R., McGuffin, P., Tennant, C., Hurry, J. (1989). The risk of minor depression before age 65: results from a community survey. Psychological Medicine, 19, 393-400.
- Bebbington, P.E., Brugha, T., MacCarthy, B., Potter, J., Sturt, E., Wykes, T., Katz, R., McGuffin, P. (1988). The Camberwell collaborative depression study. I Depressed probands: Adversity and the form of depression. British Journal of Psychiatry, 152, 754-765.
- Bebbington, P.E., Sturt, E., Tennant, C., Hurry, J. (1984). Misfortune and resilience: a community study of women. Psychological Medicine, 14, 347-363.
- Bebbington P., Tansella, M. (1989). Gender, marital status and treated affective disorders in South Verona: A case register study. Journal of Affective Disorders, 17, 83-91.
- Beck, A.T., Ward, C.M., Mendelson, M., Mock, J.E. and E., Waugh, J.K. (1961) An inventory for measuring depression. Archives of General Psychiatry, 4, 561-671.
- Belle, D. (1980). Who uses Mental Health Facilities? in The Mental Health of Women, eds. M. Guttentag, S. Salasin and D. Belle. New York: Academic Press.
- Bem, S.L. (1974). The measurement of psychological androgyny. Journal of Consulting and Clinical Psychology, 42, 155-162.
- Berah, E.F. (1983). Sex Differences in Psychiatry Morbidity: An Analysis of Victorian Data. Australian and New Zealand Journal of Psychiatry, 17, 266-173.



- Berman, E.M. and Lief, H.I. (1975) Marital therapy from a psychiatric perspective: An overview. American Journal of Psychiatry, 132, 583-592.
- Birtchnell, J. (1984). Dependence and its relationship to depression. British Journal of Medical Psychology, 57, 215-225.
- Birtchnell, J. (1985). The relationship between scores on Ryle's Marital Patterns Test and independent ratings of marital quality. British Journal of Psychiatry, 146, 638-644.
- Blacker, C.V.R., Clare, A.W. (1987). Depressive Disorders in Primary Care. British Journal of Psychiatry, 150, 737-751.
- Bland, R.C. (1988). Investigations of the prevalence of psychiatric disorders. Acta Psychiatrica Scandinavica, 77, 7-16.
- Bland, R.C., Orn, H., Newman, S.C. (1988b). Lifetime prevalence of psychiatric disorders in Edmonton. Acta Psychiatrica Scandinavica, 77, 24-32.
- Blatt, S.J., D'Affliti, J.P. & Quinlan, D.M. (1975). Depressive Experiences Questionnaire. Unpublished report, Department of Psychology, Yale University.
- Blazer, D., George, L.K., Landerman, R., Pennybacker, M., Melville, M.L., Woodbury, M., Manton, K.G., Jordan, K., Locke, B. (1985). Psychiatric Disorders, A Rural/Urban Comparison. Archives of General Psychiatry, 42, 651-656.
- Blazer, D., Swartz, M., Woodbury, M., Manton, K.G., Hughes, D., George, L.K. (1988). Depressive Symptoms and Depressive Diagnosis in a Community Population. Archives of General Psychiatry, 45, 1078-1084.
- Blouin, A.G., Perez, E.L., Blouin, J.H. (1988). Computerized Administration of the Diagnostic Interview Schedule. Psychiatry Research, 23, 335-344.
- Blumenthal, M.D. (1975). Measuring depressive symptomatology in a general population. Archives of General Psychiatry, 32, 971-978.
- Bolton, W., Oatley, K. (1987). A longitudinal study of social support and depression in unemployed men. Psychological Medicine, 17, 453-460.

- Bowlby, J. (1973). Attachment and Loss. Vol. 2: Separation, Anxiety and Anger, 366-410. London: Hogarth Press.
- Boyce, P., & Parker, G.B. (1989). Development of a Scale to Measure Interpersonal Sensitivity. Australian and New Zealand Journal of Psychiatry, 23, 341-351.
- Boyce, P., Hadzi-Pavlovic, D., Parker, G., Brodaty, G., Hickie, I., Mitchell, P., Wilhelm, K. (1989) Depressive type and state effects on personality measures. Acta Psychiatrica Scandinavica 81, 197-200.
- Boyce, P., Parker, G.B., Barnett, B., Cooney, M., Smith, J. (1989). Personality as a Vulnerability Factor to Depression. British Journal of Psychiatry (in press).
- Boyd, J.H. & Weissman, M.M. (1981). Epidemiology of affective disorders: a reexamination and future directions. Archives of General Psychiatry 38, 1039-1046.
- Boyd, J.H. & Weissman, M.M. (1982). Epidemiology. in Handbook of Affective Disorders, ed. Paykel, E.S. New York: Churchill Livingstone.
- Boyd, J.H., Burke, J.D., Gruenberg, E., Holzer, E., Rae, D.S., George, L.K., Karno, M., R. Stoltzman, R., McEvoy, L., Nestadt, G. (1984). Exclusion Criteria of DSM-III. A study of Co-occurrence of Hierarchy-Free Syndromes. Archives of General Psychiatry, 41, 893-989.
- Breier, A., Charney, D.S., Heninger, G.R. (1985). The diagnostic validity of anxiety disorders and their relationship to depressive illness. American Journal of Psychiatry, 142, 787-797.
- Briscoe, M. (1982). Sex differences in psychological well-being. Psychological Medicine, Monograph Supplement 1.
- Briscoe, M.E., Vazquez-Barquero, J.L., Williams, P., Diez-Manrique, J.F., Pena, C. (1989). Sex Differences in the Differentiation of Psychiatric Symptomatology. British Journal of Psychiatry, 154, 364-367.
- Bromet, E.J., Dunn, L.O., Connell, M.M., Dew, M.A., Schulberg, H.C. (1986). Long-term reliability of diagnosing lifetime major depression in a community sample. Archives of General Psychiatry, 43, 435-440.

- Brown, C.H., Shapiro, S., Kramer, M. & Gruenberg, E.M. (1985). Comparison of the lay Diagnostic Interview Schedule and a standardized psychiatric diagnosis: Experience in Eastern Baltimore. Archives of General Psychiatry, 42, 667-675.
- Brown, G. and Harris T. (1978). Social Origins of Depression. A Study of Psychiatric Disorder in Women. London: Tavistock Publications.
- Brown, G.W., Adler, Z., Bifulco, A., (1988). Life events, difficulties and recovery from chronic depression. British Journal of Psychiatry, 152, 487-498.
- Brown, G.W., Andrews, B., Harris, T., Adler, Z., Bridge, L., (1986). Social support, self-esteem and depression. Psychological Medicine, 16, 813-831.
- Brown, G.W., Bifulco, A., (1990). Motherhood, employment and the development of depression: a replication of a finding. British Journal of Psychiatry, 156, 169-179.
- Brown, G.W., Bifulco, A., and Harris, T.O. (1987). Life events, vulnerability and onset of depression: some refinements. British Journal of Psychiatry 1, 150, 30-42.
- Brown, G.W., Bifulco, A., Harris, T. O., (1987). Life events, vulnerability and onset of depression: some refinements. British Journal of Psychiatry, 150, 30-42.
- Brown, G.W., Craig, T.K.J., Harris, T.O. (1985). Depression: distress or disease? Some epidemiological considerations. British Journal of Psychiatry, 147, 612-622.
- Brown, G.W., Prudo, R. (1981) Psychiatric disorder in a rural and an urban population: 1 Aetiology of depression. Psychological Medicine 11, 581-599.
- Brugha, T.S., Bebbington, P.E., MacCarthy, B., Wykes, T., Sturt, E., Potter, J. (1990). Gender, social support and recovery from depressive disorders: a prospective clinical study. Psychological Medicine, 20, 147-156.
- Bryson, S.E. & Pilon, D.J. (1984). Sex differences in depression and the method of administering the Beck Depression Inventory. Journal of Clinical Psychology, 40, 529-534.
- Burke, J.D. (1986). Diagnostic Categorization by the Diagnostic Interview Schedule (DIS): A Comparison with Other Methods of Assessment. In Mental Disorders in the Community, ed. Barrett, J., Rose, R. New York: Guildford.

- Burnam, M.A., Hough, R.I., Escobar, J.I., Karno, M., Timbers, D.M., Telles, C.A., Locke, B.Z. (1987). Six-Month Prevalence of Specific Psychiatric Disorders Among Mexican Americans and Non-Hispanic Whites in Los Angeles. Archives of General Psychiatry, 44, 687-694.
- Burvill, P.W. (1987). Review: An appraisal of the NIMH epidemiologic catchment area program. Australian and New Zealand Journal of Psychiatry, 21, 175-184.
- Canino, G.J., Bird, H.R., Shrout, P.E., Rubio-Stipec, M., Bravo, M., Martinez, R., Sesman, M., Guevara, L.M. (1987a). The prevalence of Specific Psychiatric Disorders in Puerto Rico. Archives of General Psychiatry, 44, 727-735.
- Canino, G.J., Rubio-Stipec, M., Shrout, P., Bravo, M., Stolberg, R., Bird, H.R. (1987b). Sex Differences and depression in Puerto Rico. Psychology of Women Quarterly, 11, 443-459.
- Carney, M.W.P., Roth, M., Garside, R.F. (1965). The diagnosis of depressive syndromes and the prediction of ECT response. British Journal of Psychiatry, 111, 659-674.
- Carroll, B.J. (1989). Diagnostic Validity and Laboratory Studies: Rules of the Game. Validity of Psychiatric Diagnosis. New York: Raven Press.
- Cassel, J. (1976). The contribution of the social environment to host resistance. American Journal of Epidemiology, 104, 107-123.
- Charney E.A., Weissman, M.M. (1988). Epidemiology of Depression and Manic Syndromes. In Depression and Mania. New York: Elsevier.
- Christie, K.A., Burke, J.D., Regier, D.A., Rae, D.S., Boyd, J.H., Locke, B.Z. (1988). Epidemiologic Evidence for Early Onset of Mental Disorders and Higher Risk of Drug abuse in Young Adults. American Journal of Psychiatry, 145:8, 971-975.
- Clancy, K. & Gove, W. (1974) Sex differences in mental illness: and analysis of response bias in self-reports. American Journal of Sociology, 80, 205-216.
- Cleary, P.D., Mechanic, D. (1983). Sex Differences in Psychological Distress among Married People. Journal of Health and Social Behavior, 24, 111-121.
- Clinebell, H.J. and Clinebell, C.M. (1970). The Intimate Marriage. New York: Harper and Row.

- Cobb, J.P., McDonald, R., Marks, I. and Stern, R. (1980). Marital versus exposure therapy - Psychological treatment of co-existing marital and phobic-obsessive problems. Behavior Analysis and Modification, 4, 3-16.
- Cochrane, R., Stopes-Roe, M. (1981). Women, Marriage, Employment and Mental Health. British Journal of Psychiatry, 139, 373-381.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37-46.
- Cohen, J. (1977). Statistical Power for the Behavioral Sciences, third ed. Academic Press: New York.
- Collins, J., Kreitman, N., Nelson, B and Troop, J. (1971). Neurosis and marital interaction III. Family roles and functions. British Journal of Psychiatry, 119, 233-242.
- Comstock, G.W., Helsing, K.J. (1976). Symptoms of depression in two communities. Psychological Medicine, 6, 551-563.
- Congalton, A.A. (1969) Status ranking list of occupations in Australia, Appendix B to Status and Prestige in Australia, Melbourne: F.W. Cheshire.
- Cooper, J.E., Copeland, J.R., Brown, G.W., Harris, T., Gourlay, J. (1977). Further studies on Interviewer Training and Inter-rater Reliability of the Present State Examination (PSE). Psychological Medicine, 7, 517-523.
- Cooper, J.E., Kendell, R.E., Gurland, B.J., Sharpe, L., Copeland, J.R.M., Simon, R. (1972). Psychiatric diagnosis in New York and London. Maudsley Monograph no. 20. London: Oxford University Press.
- Copeland, J. (1981). What is a 'case'? A case for what? In What is a case? The Problem of Definition in Psychiatric Community Surveys, eds. J.K. Wing, P. Bebbington, L.N. Robins, London: Grant, McIntyre.
- Costello, C.G. & Comrey, A.L. (1967). Scales for measuring depression and anxiety. Journal of Psychology, 66, 303-313.
- Costello, C.G. (1982) Social factors associated with depression: a retrospective community study. Psychological Medicine, 12, 329-339.
- Cox, J.L. (1990). Childbirth as a life event: Sociocultural Aspects of postnatal depression. Acta Psychiatrica Scandinavica, 79, 75-83.

- Cox, J.L., Connor, Y., Kendell, R.E. (1982). Prospective study of the psychiatric disorders of childbirth. British Journal of Psychiatry, 140, 111-117.
- Cox, J.L., Rooney, A., Thomas, P.E., Wrate, R.W. (1984). How accurately do mothers recall postnatal depression? Further data from a 3-year follow-up study. Journal of Psychosomatic Obstetrics and Gynaecology, 3, 185-189.
- Coyne, J.C. (1987). Depression, Biology, Marriage and Marital Therapy. Journal of Marital and Family Therapy, 4, 393-407.
- Coyne, J. (1985) Towards an Interactional Description of Depression. In Essential Papers on Depression. New York: New York University Press.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.
- Crowe, M.T. (1978). Conjoint marital therapy: A controlled outcome study. Psychological Medicine, 8, 623-636.
- Crowell, B.A., George, L.K., Blazer, D., Landerman, R. (1986). Psychosocial risk factors and Urban/Rural differences in the prevalence of major depression. British Journal of Psychiatry, 149, 307-314.
- Dean, C., Surtees, P.G., Sashidharan, S.P. (1983). Comparison of Research Diagnostic Systems in an Edinburgh Community Sample. British Journal of Psychiatry, 142, 247-256.
- Department of Commonwealth Services and Health (1990). Tobacco in Australia, a summary of related statistics. Canberra: Australian Government Publishing Service.
- Dewsbury, R. (1986) Teachers learn how to live with stress. The Sydney Morning Herald, 27th August, 1986.
- Dorsch, M.M., Roder, D.M. (1983). A comparison of Australian suicide rates in 1969-1973 and 1976-1980. Australian and New Zealand Journal of Psychiatry, 17, 254-257.
- Eaton, W.W., & Kessler, L.B. (1981). Rates of symptoms of depression in a national sample. American Journal of Epidemiology, 114, 528-538.

- Eaton, W.W., Holzer, C.E., Von Korff, M., Anthony, J.C., Helzer, J.E., George, L., Burnam, A., Boyd, J.H., Kessler, L.G., Locke, B.Z. (1984). The design of the epidemiologic catchment area surveys. The control and measurement of error. Archives of General Psychiatry, 41, 942-948.
- Eaton, W.W., Regier, D.A., Locke, B.Z., Taube, C.A. (1981). The NIMH epidemiologic catchment area program. In What is a case? The Problem of Definition in Psychiatric Community Surveys, eds. J.K. Wing, P. Bebbington, L.N. Robins, London: Grant, McIntyre.
- Eaton, W.W., Ritter, C. (1988). Distinguishing anxiety and depression with field survey data. Psychological Medicine, 18, 155-166.
- Endicott, J., Spitzer, R.L. (1978). A diagnostic interview: The schedule for affective disorders and schizophrenia. Archives of General Psychiatry, 35, 837-844.
- Ensel, W.M. (1982). The role of age in the relationship of gender and marital status to depression. Journal of Nervous and Mental Disease, 170, 536-543.
- Erdman, H.P., Klein, M.H., Greist, J.H., Bass, S.M., Bires, J.K., Machtinger, P.E. (1987). A Comparison of the Diagnostic Interview Schedule and Clinical Diagnosis. American Journal of Psychiatry, 144, 1477-1480.
- Essen-Moller, E., Larsson, H., Uddenberg, C., White, G. (1956). Individual traits and morbidity in a Swedish rural population. Acta Psychiatrica et Neurologica Scandinavica, Supplement 100.
- Eysenck, H.J. & Eysenck, S.B. (1964) Manual of the Eysenck Personality Inventory. London: Hodder and Stoughton.
- Farmer, A.E., Katz, R., McGuffin, P., Bebbington, P. (1987). A Comparison between the Present State Examination and the Composite International Diagnostic Interview. Archives of General Psychiatry, 44, 1064-1068.
- Feichner, J.P., Robins, E., Guze, S.B., Woodruff, R.A., Winokur, G., Munoz, R. (1972). Diagnostic Criteria for Use in Psychiatric Research. Archives of General Psychiatry, 26, 57-63.
- Fieve, R.R., Go, R., Dunner, D.L., Elston, R. (1984). Search for biological/genetic markers in a long-term epidemiological and morbid risk study of affective disorders. Journal of Psychiatric Research, 18-425-445.

- Finlay-Jones, R. (1986). Factors in the teaching environment associated with severe psychological distress among school teachers. Australian and New Zealand Journal of Psychiatry, 20, 304-313.
- Finlay-Jones, R., Brown, G.W., Duncan-Jones, P., Harris, T., Murphy, E., Prudo, R. (1980). Depression and Anxiety in the Community: Replicating the Diagnosis of a Case. Psychological Medicine, 10, 445-454.
- Folstein, M.F., Romanoski, A., Nestadt G., Chahal, R., Merchant, A., Shapiro, S., Kramer, M., Anthony, J., Gruenberg, E.M. & McHugh, P.R. (1985). Brief report on the clinical appraisal of the Diagnostic Interview Schedule carried out at the Johns Hopkins site of the Epidemiological Catchment Area programme of the NIMH. Psychological Medicine, 15, 809-814.
- Ford, M.R., Widiger, T.A. (1989). Sex Bias in the Diagnosis of Histrionic and Antisocial Personality Disorders. Journal of Consulting and Clinical Psychology, 57, 301-305.
- Frank, E., Carpenter, L.L., Kupfer, D.J. (1988). Sex differences in recurrent depression: Are they that significant? American Journal of Psychiatry, 145:1, 41-45.
- Funabiki, D., Bologna, N.D., Pepping, M., Fitzgerald, K.C. (1980). Revisiting sex differences in the expression of depression. Journal of Abnormal Psychology, 89, 194-202.
- Galloway, D. Panckhurst, K.B. Boswell, C. & Green, K. (1984). Mental Health, Absences from Work, Stress & Satisfaction in a Sample of New Zealand Primary School Teachers. Australian & New Zealand Journal of Psychiatry, 18. (359-363).
- George, L.K., Blazer, D.G., Hughes, D.C., Fowler, N. (1989). Social support and the outcome of major depression. British Journal of Psychiatry, 154, 478-485.
- Gershon, E.S., & Bunney, W.E. (1976). The question of X-linkage in bipolar manic-depressive illness. Journal of Psychiatric Research, 13, 99-117.
- Giles, D.E., Biggs, M.M., Roffwarg, H.P., Orsulak, P.J., Rush, A.J. (1987). Secondary depression: a comparison among subtypes. Journal of Affective Disorders, 12, 251-258.
- Gitlin, M.J., & Pasnau, R.O. (1989). Psychiatric Syndromes Linked to Reproductive Function in Women: A Review of Current Knowledge. American Journal of Psychiatry, 146, 1413-1422.



- Goldberg, D.P. (1972). The detection of psychiatric illness by questionnaire (pp 139-140). Institute of Psychiatry, Maudsley Monographs, No. 21, London: Oxford University Press.
- Goldberg, D.P., Cooper, B., Eastwood, M.R., Kedward, H.B., Shepherd, M. (1970). A standardized psychiatric interview for use in community surveys. British Journal of Preventive and Social Medicine, 24, 18.
- Gonzales, L.R., Lewinsohn, P.M., Clarke, G.N. (1985). Longitudinal Follow-up of Unipolar Depressives: An Investigation of Predictors of Relapse. Journal of Consulting and Clinical Psychology, 4, 461-469.
- Gotlib, I.H. Mount, J.H. Cordy, N.I. Whiffen, V.E. (1988). Depression & Perceptions of Early Parenting: a Longitudinal Investigation. British Journal of Psychiatry. 152, 24-27.
- Gotlib, I.H., Whiffen, V.E. (1989). Depression and marital functioning: An examination of specificity and gender differences. Journal of Abnormal Psychology, 98, 23-30.
- Gove, W.R. (1984). Gender Differences in Mental and Physical Illness: The Effects of Fixed Roles and Nurturant Roles. Social Science and Medicine, 19, 77-91.
- Gove, W.R., Hughes, M. (1979). Possible causes of the apparent sex differences in physical health: an empirical investigation. American Sociological Review, 44, 126-146.
- Gove, W.R., Hughes, M., Style, C.B. (1983). Does Marriage have Positive Effects on the Psychological Well-Being of the Individual? Journal of Health and Social Behavior, 24, 122-131.
- Grove, W.M., Andreasen N.C., McDonald-Scott, P., Keller, M.B., Shapiro, R.W. (1981). Reliability Studies of Psychiatric Diagnosis. Archives of General Psychiatry, 38, 408-413.
- Gurland, B.J., Yorkston, N.J., Stone, A.R., Frank, J.D. & Fliess, L. (1972). The structured and scaled interview to assess maladjustment (SSIAM). I Description, rationale and development. Archives of General Psychiatry, 27, 259-263.
- Guze, S.B., Woodruff, R.A., Clayton, P.J. (1971). 'Secondary' affective disorder: A study of 95 cases. Psychological Medicine, 1, 426-428.

- Hafner, R.J. (1986). Marriage and mental illness: A sex Roles Perspective. The Guildford Family Therapy Series. ed. A.S. Gurman. New York: The Guildford Press.
- Hafner, R.J. (1989). Health Differences Between Married Men and Women: The Contribution of Sex-Role Stereotyping. A.N.Z. Journal of Family Therapy, 1, 13-19.
- Hagnell, O. (1966). The Lundby study on psychiatric morbidity (Sweden), in Prospective longitudinal research: an empirical basis for the primary prevention of psychosocial disorders, eds. S.A. Mednick and A.E. Baert. New York: Oxford University Press.
- Hagnell, O., Lanke, J., Rorsman, B. & Ojesjo, L. (1982). Are we entering an age of melancholy? Depressive illnesses in a prospective epidemiologic study over 25 years - the Lundby study, Sweden. Psychological Medicine, 12, 279-289.
- Halbreich, U., Vital-Herne, J., Goldstein, S., Zander, K. (1984). Sex Differences in Biological Factors Putatively Related to Depression. Journal of Affective Disorders, 7, 223-233.
- Hallstrom, T. (1987). Major depression, parental mental disorder and early family relationships. Acta Psychiatrica Scandinavica. 75, 259-263.
- Hallstrom, T. Samuelsson, S. (1985). Mental Health in the climacteric: the longitudinal study of women in Gothenburg. Acta Obstetric Gynecological Scandinavica, 130, 13-18.
- Hammen, C.L. and Padesky, C.A. (1977). Sex differences in the expression of depressive responses on the Beck Depression Inventory. Journal of Abnormal Psychology, 86, 609-614.
- Harris, T.O. and Brown, G.W. (1985). Interpreting data in aetiological studies of affective disorder: Some pitfalls and ambiguities. British Journal of Psychiatry, 147, 5-15.
- Harris, T.O., (1988). Psychosocial vulnerability to depression: the biographical perspective of the Bedford College studies. Handbook of Social Psychiatry, eds. S. Henderson, G. Burrows. Canberra: Elsevier.
- Harris, T.O., Brown, G.W., (1985). Interpreting data in aetiological studies of affective disorder: some pitfalls and ambiguities. British Journal of Psychiatry, 147, 5-15.

- Hasin, D., Link, B. (1988). Age and Recognition of Depression: Implications for a Cohort effect in Major Depression. Psychological Medicine, 18, 683-688.
- Heins, T.J., Yelland, J.H. (1981). Validity studies on the Ryle Marital Patterns Test. British Journal of Medical Psychology, 54, 51-58.
- Helgason, T. (1964). Epidemiology of mental disorders in Iceland. Acta Psychiatrica Scandinavica, 173, 1-258.
- Helgason, T., (1986). Expectancy and Outcome of Mental Disorders in Iceland In Community Surveys of Psychiatric Disorders. New Jersey: Rutgers University Press.
- Helzer, J.E., Robins, L.N., Croughan, J.L., Welner, A. (1981). Renard diagnostic interview. Its reliability and procedural validity with physicians and lay interviewers. Archives of General Psychiatry, 38, 393-398.
- Helzer, J.E., Robins, L.N., McEvoy, L.T., Spitznagel, E.L., Stoltzman, R.K., Farmer, A., Brockington, I.F. (1985). A comparison of clinical and Diagnostic Interview Schedule diagnoses. Archives of General Psychiatry, 42, 657-666.
- Helzer, J.E., Spitznagel, E.L., McEvoy, L., (1987). The Predictive Validity of Lay Diagnostic Interview Schedule Diagnosis in the General Population: A Comparison with Physician Examiners. Archives of General Psychiatry, 44, 1069-1077.
- Henderson, S., Byrne, D.G., and Duncan-Jones, P. (1981) Neurosis and The Social Environment. Sydney: Academic Press.
- Henderson, S., Duncan-Jones, P., Byrne, D.G., Scott, R., Adcock, S. (1979). Psychiatric Disorder in Canberra: A Standardized Study of Prevalence. Acta Psychiatrica Scandinavica, 60, 355-374.
- Hesselbrock, V., Stabenau, J., Hesselbrock, M., Mirkin, P., Meyer, R. (1982). A Comparison of Two Interview Schedules. Archives of General Psychiatry, 39, 674-677.
- Hillier, W., Zaudig, M., Bose, M.V. (1989). The overlap between depression and anxiety on different levels of psychopathology. Journal of Affective Disorders, 16, 223-231.

- Hinchcliffe, M., Vaughan, P.W. Hooper, D., Roberts, F.J. (1978). The melancholy marriage: An inquiry into the interaction of depression. British Journal of Medical Psychology, 51, 1-13.
- Hinde, R.A. (1979). Towards Understanding Relationships. London: Academic Press.
- Hirschfeld, R.M.A., Klerman, G.L., Andreasen, N.C., Clayton, P.J., Keller, M.B. (1986). Psycho-Social Predictors of Chronicity in Depressed Patients. British Journal of Psychiatry, 148, 648-654.
- Hirschfeld, R.M.A., Klerman, G.L., Clayton, P.J., Keller, M.B., Andreasen, N.C. (1984). Personality and Gender-Related Differences in Depression. Journal of Affective Disorders, 7, 211-221.
- Hirschfeld, R.M.A., Klerman, G.L., Gough, H.G., Barrett, J., Kordins, J. & Chodoff, P. (1977). A measure of interpersonal dependency. Journal of Personal Assessment, 4, 610-618.
- Hodiamont, P., Peer, N., Syben, N. (1987). Epidemiological aspects of psychiatric disorder in a Dutch health area. Psychological Medicine, 17, 495-505.
- Holden, C. (1986). Proposed New Psychiatric Diagnosis Raise Charges of Gender Bias. Science, 231, 327-328.
- Holmes, T.H., and Rahe, R.H. (1967). The social readjustment rating scale. Journal of Psychosomatic Research, 11, 213-218.
- Holmström, R., Jussila, L., Vauhkonen, K. (1987). Assessment of mental health and illness considered in the light of a 13 year longitudinal study. Acta Psychiatrica Scandinavica, 75, Supplement 33, 7-63.
- Hunt, G.E., Johnson, G.F.S., Caterson, I.D. (1989). The effect of age on cortisol and plasma dexamethasone concentrations in depressed patients and controls. Journal of Affective Disorders, 17, 21-32.
- Hurry, J., Bebbington, P.E., Tennant, C. (1987). Psychiatric symptoms, social disablement and illness behaviour. Australian and New Zealand Journal of Psychiatry, 21, 68-73.
- Ingham, J.G., Kreitman, N.B., Miller, P. McC., Sashidharan, S.P., Surtees, P.G. (1986). Self-Esteem, Vulnerability and Psychiatric Disorder in the Community. British Journal of Psychiatry, 148, 375-385.

- Ingham, J.G., Kreitman, N.B., Miller, P. McC., Sashidharan, Surtees, P.G., (1987). Self-appraisal, anxiety and depression in women: a prospective enquiry. British Journal of Psychiatry, 151, 643-651.
- Jacobson, D., (1987). Models of stress and meanings of unemployment; reactions to job loss among technical professionals. Social Science and Medicine, 24, 13-21.
- Jampala, V., Sierles, F.S., Taylor, M.A. (1988). The use of DSM-III in the United States: A case of not going by the book. Psychiatry and Behavioural Sciences, 22, 39-47.
- Jenkins, R. (1985). Sex differences in minor psychiatric morbidity. Psychological Medicine, Monograph Supplement 7.
- Jorm, A.F. (1987). Sex and age differences in depression: A quantitative synthesis of published research. Australian and New Zealand Journal of Psychiatry 21, 46-53.
- Jorm, A.F., Duncan-Jones, P., Scott, R. (1989). An analysis of the re-test artefact in longitudinal studies of psychiatric symptoms and personality. Psychological Medicine, 19, 487-493.
- Kandel, D.B., Davies, M. (1982). Epidemiology of Depressive Mood in Adolescents. Archives of General Psychiatry, 39, 1205-1212.
- Kaplan, G.A., Roberts, R.E., Camacho, T.C., Coyne, J.C. (1987). Psychosocial Predictors of Depression: Prospective Evidence from the Human Population Laboratory Studies. American Journal of Epidemiology. 7, 786-792.
- Kaplan, M. (1983). A woman's view of DSM-III. American Psychologist, 7, 786-792.
- Karno, M., Hough, R.L., Burnam, M.A., Escobar, J.I., Timbers, D.M., Santana, F., Boyd, J.H. (1987). Lifetime Prevalence of Specific Psychiatric Disorders among Mexican Americans and Non-Hispanic Whites in Los Angeles. Archives of General Psychiatry, 44, 695-701.
- Katz, M.M., Klerman, G.L. (1979). Introduction: overview of the clinical studies program. American Journal of Psychiatry, 136, 1-10.
- Katz, R., McGuffin, P. (1987). Neuroticism in familial depression. Psychological Medicine, 17, 155-161.

- Keller, M.B., Lavori, W., Friedman, B., Nielsen, E., Endicott, J., McDonald-Scott, P., Andreasen, N.C., (1987). The longitudinal interval follow-up evaluation. Arch. Gen. Psychiatry, 44, 540-548.
- Kendell, R.E. (1969). The continuum model of depressive illness. Proceedings of the Royal Society of Medicine, 62, 335-339.
- Kendell, R.E. (1976). The classification of depressions: A review of contemporary confusion. British Journal of Psychiatry, 129, 15-28.
- Kendell, R.E. (1988). What is a case? Archives of General Psychiatry, 45, 374-376.
- Kendell, R.E. (1989). Clinical validity. Psychological Medicine, 19, 45-55.
- Kendell, R.E., Chalmers, J.C., Platz, C. (1987). Epidemiology of puerperal psychoses. British Journal of Psychiatry, 150, 662-673.
- Kessler, R.C., McLeod, J.D., (1984). Sex Differences in Vulnerability to Undesirable Life Events. American Sociological Review, 49, 620-631.
- Kiloh, L.G., Andrews, G., Neilson, M., Bianchi, G.N. (1972). The relationship of the syndromes called endogenous and neurotic depression. British Journal of Psychiatry, 121, 183-196.
- Kiloh, L.G., Garside, R.F. (1963). The independence of neurotic depression and endogenous depression. British Journal of Psychiatry, 109, 451-463.
- King, D.A., & Bochwald, A.M. (1982). Sex differences in sub-clinical depression: Administration of the Beck Depressive Inventory in public and private disclosure situations. Journal of Personality and Social Psychology, 42, 963-969.
- Kirk, S. & Cook, D. (1990). Blackboard Grumble. The Sydney Morning Herald, 31st January 1990.
- Klerman, G.L. (1980). Overview of Affective Disorders. Chapter 18.1. in Comprehensive Textbook of Psychiatry, eds. Kaplan, I.H., Freedman, A.M., Sadock, B.J. 4th Edition. Baltimore: Williams and Wilkins.
- Klerman, G.L. (1985a). Diagnosis of Psychiatric Disorders in Epidemiologic Field Studies. Archives of General Psychiatry, 42, 723-724.

- Klerman, G.L. (1985b). The Current Age of Youthful Melancholia: Evidence of Increase in Depression among Adolescents and Young Adults. British Journal of Psychiatry, 152, 4-14.
- Klerman, G.L., Endicott, J., Spitzer, R., Hirschfeld, R.M.A. (1979). Neurotic depressions: a systematic analysis of multiple criteria and meanings. American Journal of Psychiatry, 136. 57-61.
- Klerman, G.L., Lavori, P.W., Rice, J., Reich, T., Endicott, J., Andreasen, N.C., Keller, M.B., Hirschfeld, R.M.A. (1985). Birth-cohort trends in rates of major depressive disorder among relatives of patients with affective disorder. Archives of General Psychiatry, 42, 689-693.
- Kocsis, J.H., Frances, A.J. (1987). A Critical Discussion of DSM-III Dysthymic Disorder. American Journal of Psychiatry, 144, 1534-1542.
- Koslow, S.H., Gaist, P.A. (1987). The Measurement of Neurotransmitters. In The Measurement of Depression, eds. A.J. Marsella, R.M.A. Hirschfeld, M.M. Katz. New York: Guildford.
- Kraupl-Taylor, F. (1972). A Logical Analysis of the Medico-Psychological Concept of Disease, Part 2. Psychological Medicine, 2, 7-16.
- Kreitman, N. (1964). The patient's spouse. British Journal of Psychiatry, 110, 159-173.
- Kreitman, N., Collins, J., Nelson, B. and Troop, J. (1971). Neurosis and marital interaction. IV Manifest psychological interaction. British Journal of Psychiatry, 117, 33-46.
- Krupinski, J., Mackenzie, A. (1979). The Health and Social Survey of the North West Region of Melbourne. Melbourne: Mental Health Division, Special Publication No. 7.
- Krupinski, J., Stoller, A., Baikie, A.G., Graves, J.E. (1970). A Community Health Survey of the Rural Town of Heyfield, Victoria, Australia. Melbourne: Mental Health Authority, Special Publication No. 1.
- Kyriacou, C. & Sutcliffe, (1977). Teacher Stress : a review. Educational Review, 29, 299-309.
- Leckman, J.F., Sholomskas, D., Thompson, D., Belanger, A., & Weissman, M.M. (1982). Best estimate of lifetime prevalence diagnosis: A methodological study. Archives of General Psychiatry, 39, 879-883.

- Leighton, A.H. (1959). My name is Legion. The Stirling County Study of Psychiatric Disorder and Sociocultural Environment, Vol I. New York: Basic Books.
- Leighton, A.H. (1963). The Initial Frame of Reference of the Stirling County Study: Main Questions Asked & Reasons for Them. In Mental Disorders in the Community, ed. Barrett, J., Rose, R. New York: Guildford.
- Lemkau, P.V. (1955). Mental Hygiene in Public Health. New York: McGraw Hill.
- Lewinsohn, P.M., Duncan, E.M., Stanton, A.K., Hautzinger, M. (1986). Age at first onset for bipolar depression. Journal of Abnormal Psychology, 4, 378-383.
- Leonhard, K., Korff, I., Schulz, H. (1962). Temperament in families with monopolar and bipolar phasic psychoses. Psychiatric Neurology, 143, 416-434.
- Locke, H.J. and Wallace, K.L. (1959) Short marital-adjustment and prediction tests: Their reliability and validity. Marriage and Family Living, 21, 251-255.
- Loring, M. & Powell, B. (1988). Gender, Race and DSM-III: A Study of the Objectivity of Psychiatric Diagnostic Behaviour. Journal of Health and Social Behaviour, 29, 1-22.
- Lumby, C. (1989). Why our teachers quit. The Sydney Morning Herald, 1st November, 1989.
- Luria, R.E., Berry, R. (1979). Reliability and descriptive validity of PSE syndromes. Archives of General Psychiatry, 36, 1187-1195.
- Maas, J.W., Koslow, S.H., Davis, J.M., Katz, M.M., Mendels, J., Robins, E., Stokes, P.E., Bowden, C.L. (1980). Biological component of the NIMH clinical research branch collaborative program on the psychobiology of depression: I. Background and theoretical considerations. Psychological Medicine, 10, 759-776.
- Macquarie Dictionary (1982). Gen. Editor: R. Blair. Sydney: Jacaranda Press.
- Maes, M. Vandewoude, M. Maes, L. Schotte, C. and Cosyns, P. (1989). A revised interpretation of the TRH test results in female depressed patients. Journal of Affective Disorders, 16. 215-221.



- Maes, M., Vandewoude, M., Schotte, C., Maes, L., Martin, M., Block (1989). Sex-linked differences in cortisol, ACTH and prolactin responses to 5-hydroxytryptophan in healthy controls and minor and major depressed patients. Acta Psychiatrica Scandinavica, 80, 584-590.
- Mazure, C., Gershon, E.S. (1979). Blindness and Reliability in lifetime psychiatric diagnosis. Archives of General Psychiatry, 36, 521-525.
- McKinlay, J.B., McKinlay, S.M., Brambilla, D. (1987). Journal of Health and Social Behaviour, 28, 345-363.
- McLeod, J.D., Turnbull, J.E., Kessler, R.C., Abelson, J.M. (1989). Sources of discrepancy in the comparison of a lay-administered diagnostic instrument with clinical diagnosis. Psychiatry Research, 31, 145-159.
- Merikangas, K.R., Weissman, M.M., Pauls, D.L. (1985). Genetic factors in the sex ratio of major depression. Psychological Medicine, 15, 63-69.
- Miller, P. Mc., Kreitman, N.B., Ingham, J.G., Sashidharan, S.P. (1989). Self-esteem, life stress and psychiatric disorder. Journal of Affective Disorders, 17, 65-75.
- Monroe, S.M., Steiner, S.C. (1986). Social support and psychopathology: Interrelations with preexisting disorder, stress and personality. Journal of Abnormal Psychology, 1, 29-39.
- Morton, R.F. & Hebel, J.R. (1979). A Study Guide to Epidemiology and Biostatistics. Baltimore: Baltimore University Press.
- Murphy, J.M. (1986a). Trends in depression and anxiety: Men and women. Acta Psychiatrica Scandinavica, 73, 113-127.
- Murphy, J.M. (1986b). Prevalence and outcome of Depression and Anxiety disorders: Findings from the Stirling County Study. In Mental Disorders in the Community, ed. Barrett, J.E., Rose, R. New York: Guildford.
- Murphy, J.M., Neff, R.K., Sobol, A.M., Rice, J.X., Olivier, D.C. (1985). Computer diagnosis of depression and anxiety: The Stirling County Study. Psychological Medicine, 15, 99-112.
- Myers, J.K., Weissman, M.M. (1980). Use of self-report symptom scale to detect depression in a community sample. American Journal of Psychiatry, 137, 1081-1084.

- National Health and Medical Research Council. (1969). Report on a National Morbidity Survey, Parts I and II. Canberra, (1962-3).
- Newman, S.C., Bland, R.C., Orn, H., (1988). Morbidity risk of psychiatric disorders. Acta Psychiatrica Scandanavia, 77, 50-56.
- Ni Bhrölchain, M. (1979). Psychotic and neurotic depression. Some problems of method. British Journal of Psychiatry, 134, 87-93.
- Nolen-Hoeksema, S. (1987). Sex Differences in Unipolar Depression: Evidence and Theory. Psychological Bulletin, 101, 259-282.
- Norusis, M.J., (1988). Statistical Package for the Social Sciences, SPSS/PC + V2.0 Base Manual for the IBM PC/XT/AT and PS/2. New York: McGraw Hill.
- O'Connor, P. and Brown, G.W. (1984) Supportive relationships: Fact or fancy? Journal of Social and Personal Relationships, 1, 159-175.
- O'Hara, M.W., & Zekoski, E.M. (1988). Postpartum depression: a comprehensive review. in Motherhood and Mental Illness II, eds. Kumar, R., & Brockington, I.F. London: Wright.
- O'Hara, M.W., Kohout, F.J., Wallace, R.B. (1985). Depression among the Rural Elderly: A study of Prevalence and Correlates. The Journal of Nervous and Mental Disease, 173, 582-589.
- O'Hara, M.W., Neunaber, J, Zekoski, E.M. (1984). Prospective study of post-partum depression: Prevalence, course and predictive factors. Journal of Abnormal Psychology, 93, 158-171.
- Oakley-Browne, M.A., Joyce, P.R., Wells, E., Bushness, J.A., Hornblow, A.R. (1989). Christchurch psychiatric epidemiology study. Part II: Six month and other period prevalences of specific psychiatric disorders. Australian and New Zealand Journal of Psychiatry, 23, 327-340.
- Oliver, J.M. & Simmons, M.E. (1985). Affective disorders and depression as measured by the Diagnostic Interview Schedule and the Beck Depression Inventory in an unselected adult population. Journal of Clinical Psychology, 41, 469-476.
- Orley, J., & Wing, J.K. (1979). Psychiatric disorders in two African villages. Archives of General Psychiatry, 36, 513-520.

- Parker, G. (1979). Sex differences in non-clinical depression. Australian and New Zealand Journal of Psychiatry, 13, 127-132.
- Parker, G. (1983). Parental Overprotection: A Risk Factor in Psychosocial Development. New York: Grune and Stratton.
- Parker, G. (1983b). Parental 'affectionless control' as an antecedent to adult depression - a risk factor delineated. Archives of General Psychiatry, 40, 956-960.
- Parker, G. (1987). Are the lifetime prevalence estimates in the ECA study accurate? Psychological Medicine 17, 275-28.
- Parker, G. (1988). Parental style & parental loss. In Handbook of Social Psychiatry, eds. Henderson S., Burrows G. Canberra: Elsevier.
- Parker, G., Hadzi-Pavlovic, D. (1984). Modification of levels of depression in mother-bereaved women by parental and marital relationships. Psychological Medicine, 14, 125-135.
- Parker, G., Tupling, M. and Brown, L.B. (1979). A parental bonding instrument. British Journal of Medical Psychology, 52, 1-10.
- Paykel, E.A. (1974). Recent life events and clinical depression. In Life Stress and Illness, eds. E.K. Gunderson and R.H. Rahe, Springfield, Ill: Charles C. Thoms.
- Paykel, E.S., Myers, J.K., Dienes, M.N., Klerman, G.L., Lindenthal, J.J. and Pepper, M.P. (1969) Life events and depression: A controlled study. Archives of General Psychiatry, 21, 753-760.
- Pearlin, L.I., Schooler, C. (1978). The Structure of Coping. Health and Social Behaviour, 19, 2-21.
- Perris, C. (1966). A study of bipolar (manic-depressive) and unipolar recurrent depressive psychosis. Acta Psychiatrica Scandinavica, (Supplement 194), 1-89.
- Pond, D.A., Ryle, A. and Hamilton, M. (1963). Marriage and neurosis in a working class population. British Journal of Psychiatry, 109, 592-598.
- Quinton, D., Rutter, M. and Liddle, C. (1984). Institutional rearing, parenting difficulties and marital support. Psychological Medicine, 14, 107-124.

- Rabkin, J.G., Klein, D.F. The Clinical Measurement of Depressive Disorders. (1987). In The Measurement of Depression, eds. A.J. Marsella, R.M.A. Hirschfield, M.M. Katz, New York:Guildford.
- Radloff, L. (1975). Sex Differences in Depression: The effects of occupation and marital status. Sex Roles, 1, 249-265.
- Radloff, L.S. (1977). The CES-D scale: a self-report depression scale for research in the general population. Journal of Applied Psychological Measures, 1, 305-401.
- Radloff, L.S. (1980). Risk factors for depression: In The Mental Health of Women, eds. M. Guttentag, S. Salasin and D. Belle. New York:Academic Press.
- Radloff, L.S. (1985) Risk Factors for Depression: What do we Learn from Them? In Essential Papers on Depression, ed. J.C.Coyne, New York: New York University Press.
- Regier, D.A., Boyd, J.H., Burke, J.D., Rae, D.S., Myers, J.K., Kramer, M., Robins, L.N., George, L.K., Karno, M., Locke, B.Z. (1988). One-month prevalence of mental disorders in the United States. Archives of General Psychiatry, 45, 977-986.
- Regier, D.A., Hirschfeld, R.M.A., Goodwin, F.K., Burke, J.D., Lazar, J.B., Judd, L.L. (1988). The NIMH depression awareness, recognition and treatment program: structure, aims and scientific basis. American Journal of Psychiatry, 145, 11.
- Regier, D.A., Myers, J.K., Dramer, M., Robins, L.H., Blazer, D.G., Hough, R.L., Eaton, W.W. & Locke, B.Z. (1984). The NIMH Epidemiologic Catchment Area program: Historical context, major objectives and study population characteristics. Archives of General Psychiatry, 41, 934-941.
- Reich, J. (1986). The epidemiology of anxiety. Journal of Nervous and Mental Disease, 174, 129-136.
- Reich, T., Rice, J., Andreasen, N.& Clayton, P. (1980). A preliminary analysis of the segregation distribution of primary major depressive disorder. Psychiatric Bulletin, 16, 34-36.
- Renwick, P.A. & Lawler, E.E. (1978). Special report on work papers. Psychology Today 11, 53-65.

- Reynolds, I., Rizzo, C., Gallagher, H., Speedy, B., (1979). Psychosocial problems of Sydney Adults. Health Commission of N.S.W. Division of Health Services Research Medicine Referral Centre.
- Rice, J., Reich, T., Andreasen, N.C., Lavori, P.W., Endicott, J., Clayton, P.J., Keller, M.B., Hirschfeld, R.M.A., Klerman, G.L. (1984). Sex-Related Differences in Depression: Familial Evidence. Journal of Affective Disorders, 71, 199-210
- Robins, E., Guze, S.B. (1972). Classification of affective disorders. In Recent Advances in the Psychobiology of the Depressive Illnesses. eds. T.A. Williams, M.M. Katz, J.A. Shield Jr. (p283), United States Government Printing Office, Washington, D.C.
- Robins, L. N. (1989). Diagnostic Grammar and Assessment: Translating Criteria into Questions. Psychological Medicine, 19, 57-68.
- Robins, L.N. (1985). Epidemiology: Reflections on testing the validity of Psychiatric Interviews. Archives of General Psychiatry, 42, 918-924.
- Robins, L.N., Helzer, J.E. (1988). New diagnostic instruments. In Handbook of Social Psychiatry, eds. S. Henderson, G. Burrows. Sydney: Elsevier.
- Robins, L.N., Helzer, J.E., Croughan, J. & Ratcliff, K.S. (1981). National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics and validity. Archives of General Psychiatry, 38, 381-389.
- Robins, L.N., Helzer, J.E., Weissman, M.M., Orvaschel, H., Gruenberg, G., Burke, J.D. & Regier, D.A. (1984). Lifetime prevalence of specific psychiatric disorders in three sites. Archives of General Psychiatry, 41, 949-958.
- Robins, L.N., Wing, J., Wittchen, H.U., Helzer, J.E., Babor, T.F., Burke, J., Farmer, A., Jablenski, A., Pickens, R., Regier, D.A., Sartorius, N., Towle, L. (1988). The Composite International Diagnostic Interview. Archives of General Psychiatry, 45, 1069-1077.
- Robson, P.J., (1988). Self-esteem - a psychiatric view. British Journal of Psychiatry, 153, 6-15.
- Romans-Clarkson, S.E., Walton, V.A., Herbison, G.P. & Mullen, P. (1988). A study of women who refused to participate in a community survey of psychiatric disorder. Australian and New Zealand Journal of Psychiatry, 22, 19-29.

- Rosenberg, M. (1965). The Measurement of Self-Esteem. Society and the Adolescent Self-Image. Princeton, New Jersey: Princeton University Press.
- Ross, H.E., Glaser, F.B., Stiasny, S. (1988). Sex Differences in the Prevalence of Psychiatric Disorders in Patients with Alcohol and Drug Problems. British Journal of Addiction, 83, 1179-1192.
- Roth, M., Barnes, T.R.E. (1981). The classification of affective disorders: A synthesis of old and new concepts. Comprehensive Psychiatry, 22, 54-77.
- Rounsaville, B.J., Prusoff, B.A., Weissman, M.M., (1980). The course of marital disputes in depressed women: a 48-month follow-up study. Comprehensive Psychiatry, 21, 111-117.
- Russell, D. (1986). Psychiatric diagnosis and the oppression of women. Women and Therapy 5, 83-97.
- Ryle, A. (1966). A marital patterns test for use in psychiatric research. British Journal of Psychiatry, 123, 535-541.
- Sartorius, N. (1989). Recent Research activities in WHO's Mental Health Programme. Psychological Medicine, 19, 233-244.
- Schaefer, M.T. and Olsen, D.H. (1981). Assessing intimacy: The PAIR inventory. Journal of Marital and Family Therapy, 47-60.
- Scott-Heyes, G. (1982) Analysis and revision of Ryle's Marital Patterns Test. British Journal of Medical Psychology, 55, 67-75.
- Shapiro, S., Skinner, E.A., Kessler, L.G., Von Korff, M., German, P.S., Tischler, G.L., Leaf, P.J., Benham, L., Cottler, L., Regier, D.A. (1984). Utilization of health and mental health services. Three epidemiologic catchment area sites. Archives of General Psychiatry, 41, 971-978.
- Spanier, G.B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. Journal of Marriage and the Family, 38, 15-28.
- Spanier, G.B. (1979). The measurement of marital quality. Journal of Sex and Marital Therapy, 5, 288-299.

- Spitzer, R.L., Endicott, J. & Robins, E. (1981). Research Diagnostic Criteria (RDC) for a Selected Group of Functional Disorders. Updated Third Edition (6/80). Typescript obtained from New York State Psychiatric Institute (Dr Endicott).
- Spitzer, R.L., Endicott, J., Robins, E. (1978). Research Diagnostic Criteria: Rationale and reliability. Archives of General Psychiatry, 35, 773-832.
- Spitzer, R.L., Williams, J.B.W. (1983). Instruction Manual for the Structured Clinical Interview for DSM-III (SCID). New York: Biometrics Research Department, New York State Psychiatric Institute.
- Srole, L. Fischer, A.K. (1980) The Midtown Manhattan Longitudinal Study vs 'The Mental Paradise Lost' Doctrine. Arch. General Psychiatry, 37, 209-221.
- Srole, L., Langner, T.S., Michael, S.T., Kirkpatrick, P., Opler, M.K., Rennie, T.A.C., (1978). Mental health in the metropolis. The midtown Manhattan study. New York: New York University Press.
- Stavrakaki, C., Vargo, B. (1986). The relationship of anxiety and depression: A review of the Literature. British Journal of Psychiatry, 149, 7-16.
- Steele, R.E. (1978). Relationship of Race, Sex, Social Class, and Social Mobility to Depression in Normal Adults. The Journal of Social Psychology, 104, 37-47.
- Stein, G. (1982). The Maternity Blues. In Motherhood and Mental Illness, eds. I.F. Brockington & R. Kumar. London: Academic Press.
- Steinberg, F.C., Leichner, P.P., Harper, D.W. (1987). The interaction of sex-role identity and ideology conflicts with depression in women and men. Social Psychiatry, 22, 8-13.
- Stengel, E. (1959). Classification of mental disorders. Bulletin of World Health Organization, 21, 601-611.
- Sturt, E., (1981). Hierarchical patterns in the distribution of psychiatric symptoms. Psychological Medicine, 11, 783-794.
- Sturt, E., Kumakura, N., Der, G. (1984). How depressing life is - life-long morbidity risk for depressive disorder in the general population. Journal of Affective Disorders, 7, 109-122.

- Surtees, P.G. (1980). Social support, residual adversity and depressive outcome. Social Psychiatry, 15, 71-80.
- Surtees, P.G., Dean, C., Ingham, J.G., Kreitman, N.B., Miller, P. McC., Sashidharan, S.P. (1983). Psychiatric Disorder in Women from an Edinburgh Community: Associations with Demographic Factors. British Journal of Psychiatry, 142, 238-246.
- Surtees, P.G., Sashidharan, S.P., Dean, C. (1986). Affective Disorder amongst Women in the General Population: A Longitudinal Study. British Journal of Psychiatry, 148, 176-186.
- Susskind, A. (1987). Anger over promotion scheme for teachers. The Sydney Morning Herald, 2nd May 1987.
- Temkin-Greener, H., Clark, K.T. (1988). Ethnicity, gender and utilization of mental health services in a medicaid population. Social Science and Medicine, 26, 989-996.
- Tennant, C., Andrews, G. (1976). A scale to measure the stress of life events. Australian and New Zealand Journal of Psychiatry, 10, 27-32.
- Tennant, C.C. (1985). Female vulnerability to depression. Psychological Medicine, 15, 733-737.
- Torgeson, S. (1988). Neurotic Depression and DSM-III. Acta Psychiatrica Scandinavica, 76, 31-34.
- Tousignant, M., Brosseau, R., Tremblay, L. (1987). Sex basis in mental health scales: Do women tend to report less serious symptoms and confide more than men. Psychological Medicine, 17, 203-215.
- Turner, R.J., Avison, W.R. (1988). Gender and Depression: Assessing Exposure and Vulnerability to Life Events in a Chronically Strained Population. Manuscript from author.
- Uhlenhuth, E.H., Balter, M.B., Mellinger, G.D., Cisin, I.H., Clinthorne, J. (1983). Symptom checklist syndromes in the general population: Correlations with psychotherapeutic drug use. Archives of General Psychiatry, 40, 1167.
- Vaillant, G.E. (1977). Adaptation to Life. Boston: Little Brown and Co.
- Vaillant, G.E., Schnurr, P. (1988). What is a case? Archives of General Psychiatry, 45, 313-319.



- Vazquez-Barquero, J.L., Diez-Manrique, J.F., Aldama, C.P.J., Rodriguez, C.S., Arango, J.M., Mirapeix, C. (1987). A community mental health survey in Cantabria: a general description of morbidity. Psychological Medicine, 17, 227-241.
- Verbrugge, L.M. (1985). Gender and Health: An update on Hypotheses and Evidence. Journal of Health and Social Behavior, 26, 156-182.
- Veroff, J.B. (1981). The Dynamics of Help-seeking in Men and Women: A National Survey Study. Psychiatry, 44, 189-200.
- Von Zerssen, D., Weyerer, S. (1982). Sex Differences in Rates of Mental Disorders. International Journal of Mental Health, 11, 9-45.
- Warheit, G.J., Holzer, C.E., Schwab, J.J., (1973). An analysis of social class and racial differences in depressive symptomatology: A community study. Journal of Health and Social Behaviour, 14, 291-299.
- Waring, E.M. (1985). Measurement of intimacy: Conceptual and methodological issues of studying close relationships. Psychological Medicine, 15, 9-14.
- Waring, E.M. and Reddon, J.R. (1983). The measurement of intimacy in marriage: The Waring Intimacy Questionnaire. Journal of Clinical Psychology, 39, 53-57.
- Waring, E.M., McElrath, D., Lefcoe, D., Weisz, G. (1981). Dimensions of intimacy in marriage. Psychiatry, 4, 471-474.
- Waring, E.M., Tillman, M.P., Frelick, L., Russell, L. and Weisz, G. (1980). Concepts of intimacy in the general population. Journal of Nervous and Mental Disease, 168, 471-474.
- Weissman, A.N. (1980). Assessing depressogenic attitudes: A validation study. Paper presented at the annual meeting of the Eastern Psychological Association, Hartford, Connecticut.
- Weissman, M.M. & Klerman, G.L. (1977). Sex differences and the epidemiology of depression. Archives of General Psychiatry, 34, 98-111.
- Weissman, M.M. & Myers, J.K. (1978). Affective disorders in a US urban community. The use of Research Diagnostic Criteria in an epidemiological survey. Archives of General Psychiatry, 35, 1304-1311.

- Weissman, M.M. & Myers, J.K. (1978b). Rates and risks of depressive symptoms in a United States urban community. Acta Psychiatrica Scandinavia, 57, 219-231.
- Weissman, M., Pincus, C., Radding N., Lawrence, R., Siegel, R. (1973). The educated housewife: mild depression and the the search for work. American Journal Orthopsychiatry, 43, 565-573.
- Weissman, M.M., Prusoff, B.A., Klerman, G.L. (1978). Personality and th Prediction of Long-term Outcome of Depression. American Journal of Psychiatry, 135, 797-800.
- Weissman, M.M. (1987). Advances in Psychiatric Epidemiology: Rates and Risks for Major Depression. American Journal of Public Health, 77, 445-451.
- Weissman, M.M. Prusoff, B., Pincus, C. (1975). Symptom patterns in depressed patients and depressed normals. The Journal of Nervous and Mental Disease, 160, 15-23.
- Weissman, M.M., Leaf, P.J., Bruce, M.L., Florio, L. (1988). The epidemiology of dysthymia in five communities: Rates, risks, comorbidity and treatment. American Journal of Psychiatry, 145, 815-819.
- Weissman, M.M., Leaf, P.J., Holzer, C.E., Myers, J.K. & Tischler, G.L. (1984). The epidemiology of depression : An update on sex differences in rates. Journal of Affective Disorders, 7, 179-188.
- Weissman, M.M., Leaf, P.J., Tischler, G.L., Blazer, D.G., Karno, M., Bruce, M.L., Florio, L.P., (1988a). Affective Disorders in five United States communities. Psychological Medicine, 18, 141-153.
- Weissman, M.M., Myers, J.K., Ross, E.R., (1986). Community Studies in Psychiatric Epidemiology: An Introduction. In Community Surveys of Psychiatric Disorders. New Jersey: Rutgers University Press.
- Weissman, M.M., Porrenger, M., Klever, H., Ruben, H.L., Williams, D., Thompson, W.D. (1977). Symptom patterns in primary and secondary depression: A comparison of primary depressives with depressed opiate addicts, alcoholics, and schizophrenics. Archives of General Psychiatry, 34, 854-862.
- Weissman, M.M., Wickamaratne, P., Merikangas, K.R., Leckman, J.F., Prusoff, B.A., Caruso, K.A., Kidd, K.K., Gammon, D. (1984). Onset of major depression in early adulthood. Archives of General Psychiatry, 41, 1136-1143.

- Wells, J.E., Bushnell, J.A., Hornblow, A.R., Joyce, P.R., Oakley-Browne, M.A. (1989). Christchurch psychiatric epidemiology study, Part I: Methodology and lifetime prevalence for specific psychiatric disorders. Australian and New Zealand Journal of Psychiatry, 23, 315-326.
- Wells, K.B., Burnam, M.A., Leake, B., Robins, L.N. (1988). Agreement between Face-To-Face and Telephone Administered Versions of the Depression Section of the NIMH Diagnostic Interview Schedule. Journal of Psychiatric Research, 22, 207-220.
- Williams, J.B.W., Spitzer, R.L. (1982). Research diagnostic criteria and DSM-III: An annotated comparison. Archives of General Psychiatry, 39, 1283-1289.
- Williams, J.B.W., Spitzer, R.L. (1983). The issue of sex bias in DSM-III: A critique of "A woman's view of DSM-III" by Marcie Kaplan. American Psychologist, 7, 793-798.
- Williams, P., Skuse, D. (1988). Depressive thoughts in general practice attenders. Psychological Medicine, 18, 469-475.
- Wilson, P.H. (1979). Behavioural and Pharmacological Treatment of Depression. Unpublished PhD thesis, University of New South Wales.
- Wing, J.K., Cooper, J.M., Sartorius, N. (1974). The Description and Classification of Psychiatric Symptoms: An Instruction Manual for the PSE and Catego System. London: Cambridge University Press.
- Wing, J.K., Nixon, J.M., Leff, J.P. (1977). Reliability of the PSE (ninth edition) used in a population survey. Psychological Medicine, 7, 505-516.
- Wing, L. (1970). Observations on the psychiatric section of the International classification of diseases and the British glossary of mental disorders. Psychological Medicine, 1, 79.
- Winokur, G. (1985a). The validity of neurotic-reactive depression: new data and reappraisal. Archives of General Psychiatry, 42, 1116-1122.
- Winokur, G. (1985b). Controversies in Depression, or Do Clinicians Know Something After All? In Essential Papers on Depression, ed. J.C.Coyne, New York: New York University Press.
- Winokur, G., & Tanna, V.L. (1969). Possible role of X-linked dominant factor in manic-depressive disease. Diseases of the Nervous System, 30, 89-93.

- Wittchen, H.U., Burke, J.D., Semler, G., Pfister, H., Von Cranach, M., Zaudig, M. (1989). Recall and dating of psychiatric symptoms. Test-Retest reliability of time-related symptom questions in a standardized psychiatric interview. Archives of General Psychiatry, 46, 437-443.
- Wittchen, H.U., Semler, G., Von Zerssen, D. (1985). A Comparison of Two Diagnostic Methods: Clinical ICD Diagnosis vs DSM-III and Research Diagnostic Criteria using the Diagnostic Interview Schedule (Version 2). Archives of General Psychiatry, 42, 677-684.
- Wolpe, J. (1986). The positive diagnosis of neurotic depression as an etiological category. Comprehensive Psychiatry, 27, 449-460.
- World Health Organization (1973). Report of International pilot study of schizophrenia. Geneva: WHO.
- World Health Organization (1978). Mental Disorders: Glossary and guide to the classification in accordance with the Ninth Revision of the International Revision Classification of Diseases. Geneva: WHO.
- World Health Organization (1990). Draft of Chapter V, Tenth Revision of the International Revision Classification of Diseases, ICD-10. Geneva: WHO.
- Young, M., Willmott, P. (1973). The symmetrical family. Middlesex, U.K.: Penguin.
- Young, M.A., Scheftner, W.A., Fawcett, J., Klerman, G.L. (1990). Gender differences in the clinical features of unipolar major depressive disorder. The Journal of Nervous and Mental Disease, 178, 200-203.
- Zimmerman, M., Coryell, W. (1987). The inventory to diagnose depression (IDD): a self-report scale to diagnose major depressive disorder. Journal of Consulting and Clinical Psychology, 55, 55-59.
- Zimmerman, M., Coryell, W. (1988). The validity of a self-report questionnaire for diagnosing major depressive disorder. Archives of General Psychiatry, 45, 738-740.
- Zung, W.W.K. (1965). A self-rating depression scale. Archives of General Psychiatry, 12, 63-70.

---

Appendix I      Instructions, items and scoring for IBM

---

This questionnaire lists some attitudes and behaviours which people reveal in their close relationships.  
Please judge your partner's attitudes and behaviour towards you in recent times and tick the most appropriate bracket for each item.

	Very true	Moderately true	Somewhat true	Not at all
1. Is very considerate of me	(3)	(2)	(1)	(0)
2. Wants me to take his/her side in an argument	(III)	(II)	(I)	(0)
3. Wants to know exactly what I'm doing and where I am	(III)	(II)	(I)	(0)
4. Is a good companion	(3)	(2)	(1)	(0)
5. Is affectionate to me	(3)	(2)	(1)	(0)
6. Is clearly hurt if I don't accept his/her views	(III)	(II)	(I)	(0)
7. Tends to try and change me	(III)	(II)	(I)	(0)
8. Confides closely in me	(3)	(2)	(1)	(0)
9. Tends to criticise me over small issues	(III)	(II)	(I)	(0)
10. Understands my problems and worries	(3)	(2)	(1)	(0)
11. Tends to order me about	(III)	(II)	(I)	(0)
12. Insists I do exactly as I'm told	(III)	(II)	(I)	(0)
13. Is physically gentle and considerate	(3)	(2)	(1)	(0)
14. Makes me feel needed	(3)	(2)	(1)	(0)
15. Wants me to change in small ways	(III)	(II)	(I)	(0)
16. Is very loving to me	(3)	(2)	(1)	(0)
17. Seeks to dominate me	(III)	(II)	(I)	(0)
18. Is fun to be with	(3)	(2)	(1)	(0)
19. Wants to change me in big ways	(III)	(II)	(I)	(0)
20. Tends to control everything I do	(III)	(II)	(I)	(0)
21. Show his/her appreciation of me	(3)	(2)	(1)	(0)
22. Is critical of me in private	(III)	(II)	(I)	(0)
23. Is gentle and kind to me	(3)	(2)	(1)	(0)
24. Speaks to me in a warm and friendly voice	(3)	(2)	(1)	(0)

---

'CARE' scale, Arabic numerals  
'CONTROL' scale, Roman numerals

---

## **Appendix II**

**Positive and negative Life Events Scales used in the study, as described in Chapter 6.**

(A)

ID No: \_\_\_\_\_

Now examine similarly this other list of life events. Score those that occurred in the last 6 months and score their degree of associated distress.

Not distressing at all	Somewhat distressing	Fairly distressing	Very distressing	As distressing as could possibly be imagined
---------------------------	-------------------------	-----------------------	---------------------	---

1	2	3	4	5	6	7	8
0	2	4	6	8			

If YES your score

- |     |  |        |     |
|-----|--|--------|-----|
| 1.  | A major financial crisis developed   | YES/NO | ( ) |
| 2.  | You were involved in a legal action that could have damaged your reputation to a moderate or severe degree | YES/NO | ( ) |
| 3.  | You were involved in an accident that seriously threatened your life                                       | YES/NO | ( ) |
| 4.  | You were told that you were performing poorly at work  | YES/NO | ( ) |
| 5.  | You started a completely different job   | YES/NO | ( ) |
| 6.  | You were given significantly increased work responsibilities   | YES/NO | ( ) |
| 7.  | Significant problems with superiors or fellow workers developed  | YES/NO | ( ) |
| 8.  | You were down-graded or demoted at work  | YES/NO | ( ) |
| 9.  | You were dismissed from your job   | YES/NO | ( ) |
| 10. | You were unemployed (but wished to work) for more than one month   | YES/NO | ( ) |
| 11. | You were prevented from entering, or excluded from, an educational course you wished to pursue             | YES/NO | ( ) |
| 12. | You failed an important exam   | YES/NO | ( ) |
| 13. | You were told by your partner that you were no longer loved  | YES/NO | ( ) |
| 14. | Increasingly serious arguments with your partner developed   | YES/NO | ( ) |
| 15. | You separated from your partner after a breakdown in the relationship                                      | YES/NO | ( ) |
| 16. | You discovered your partner was unfaithful   | YES/NO | ( ) |
| 17. | You were formally divorced   | YES/NO | ( ) |
| 18. | You broke off your engagement to be married  | YES/NO | ( ) |
| 19. | You broke off a "steady" relationship  | YES/NO | ( ) |
| 20. | Increasingly severe arguments with your parents developed  | YES/NO | ( ) |
| 21. | Someone close to you (family member or other) developed a serious illness                                  | YES/NO | ( ) |
| 22. | You developed a serious illness, injury or operation needing hospitalization or a month or more off work   | YES/NO | ( ) |
| 23. | Your child was stillborn   | YES/NO | ( ) |
| 24. | Your partner died  | YES/NO | ( ) |
| 25. | A child of yours died  | YES/NO | ( ) |
| 26. | A close family member died (e.g. parent, brother, sister)  | YES/NO | ( ) |
| 27. | A close family friend or relative died (e.g. aunt, cousin, grandmother)                                    | YES/NO | ( ) |
| 28. | You or your partner found you were pregnant and the pregnancy was unwanted                                 | YES/NO | ( ) |
| 29. | You or your partner had an abortion for any reason   | YES/NO | ( ) |
| 30. | You or your partner had a miscarriage during a wanted pregnancy  | YES/NO | ( ) |

(A) Now please examine the following list of life events and note first, if they occurred in the last 6 months (i.e. ) and secondly, how pleasant each felt like to you using the scaling figures suggested below:

Not at all pleasing	Somewhat pleasing	Fairly pleasing	Very pleasing	As pleasing as could possibly be imagined
I	I	I	I	I
0	2	4	6	8

(You may therefore score any event that occurred as either 0, 1, 2, 3, 4, 5, 6, 7 or 8).

		If YES your score
1. You took up a new hobby or sport	YES/NO	( )
2. You moved into a new house	YES/NO	( )
3. You achieved some degree of fame	YES/NO	( )
4. You had an overseas holiday	YES/NO	( )
5. You had a holiday lasting at least a week in your own country	YES/NO	( )
6. You started a course (e.g. university or other training course) that you had been keen to commence	YES/NO	( )
7. Your financial situation improved dramatically	YES/NO	( )
8. You were able to buy something quite expensive and wanted for a long time	YES/NO	( )
9. You achieved a long-desired goal at work	YES/NO	( )
10. You were able for financial or other reasons to stop work	YES/NO	( )
11. You were told by a number of people that you were performing extremely well at work	YES/NO	( )
12. You were promoted at work	YES/NO	( )
13. You were told by someone or some group that you had helped them to a most significant extent	YES/NO	( )
14. You became aware that you had achieved a sense of fulfilment in your identity	YES/NO	( )
15. You fell in love	YES/NO	( )
16. You became aware that a partner loved you with great depth	YES/NO	( )
17. You started a "steady" relationship	YES/NO	( )
18. The relationship with your partner improved dramatically	YES/NO	( )
19. You got married	YES/NO	( )
20. You became engaged to be married	YES/NO	( )
21. You passed important exams	YES/NO	( )
22. Someone you like came to live in your house	YES/NO	( )
23. Your first child was born	YES/NO	( )
24. A child other than your first was born	YES/NO	( )
(FEMALES ONLY)		
25. You became pregnant with a wanted pregnancy	YES/NO	( )
(MALES ONLY)		
25. Your partner became pregnant with a wanted pregnancy	YES/NO	( )



## Appendix III

Scores for discriminant function analyses described in Chapter 10.

For each category (e.g. "1983 lifetime cases" in Table 1 of Appendix III), the left-hand column reports results where 'presence of a core tie' is used as a predictor variable, while the corresponding right-hand reports results with IBM 'care' and 'control' scores entered as variables. The IBM scores were entered as categorical variables, 'high care' and 'low care', 'high control' and 'low control' using the method described in Chapter 10 (see p 222).

# Appendix III

Table 1 Discriminant Function Analysis using DSM-III lifetime cases and noncases at 1983

Variable	Year	Standardised Discriminant Function Coefficients			
		1983 Lifetime Cases		1978-83 New Cases	
Sex		.09	.06	.18	.28
Maternal care	1978	-.14	-.18	.18	.18
Paternal care	1978	.60*	.87*	-.65*	-.74
Maternal protection	1978	-.48	-.47	.54	.50
Paternal protection	1978	.68*	.65*	-.71*	-.62
Neuroticism	1978	-.18	-.26	.08	.17
Trait depression	1978	-.43	-.19	.35	.16
Dependency	1978	-.27	-.12	.43	.34
Self-esteem	1978	-.09	-.01	.14	.18
BSRI (femininity) #	1983	.27	.15	-.16	-.10
Core tie	1983	.08	-	.09	-
IBM (care)	1983	-	.13	-	.14
IBM (control)	1983		.12	-	-.03
Group means	non-case	.24	.22	-.15	-.15
	case	-.78	-.76	.78	.70
Wilks' Lambda		.843	.855	.892	.905
Chi square		25.70	19.08	15.80	11.36
Significance		<.01	ns	ns	ns
Sensitivity		64.1%	62.1%	72.0%	61.9%
Specificity		73.0%	71.3%	64.3%	65.3%
Correct classification		70.9%	69.2%	65.6%	64.8%

\* p <.05

# BSRI = Bem Sex Role Inventory, femininity subscale  
 Self-esteem: higher scores reflect low self-esteem  
 Core tie : higher score indicates presence of an intimate  
 IBM care : higher score indicates low care  
 IBM control: higher score indicates high control

Table 2 Discriminant Function Analysis using DSM-III lifetime cases and noncases at 1983

Variable		Standardised Discriminant Function Coefficients			
	Year	Female	Female	Male	Male
Maternal care	1978	.05	-.19	-.36	-.17
Paternal care	1978	.54*	.73*	.53	1.03
Maternal protection	1978	-.68*	-.60	.26	.26
Paternal protection	1978	.56*	.50	.91	1.20
Neuroticism	1978	-.21	-.23	-.32	-.60
Trait depression	1978	-.36	-.32	-.01	.45
Dependency	1978	.23	.02	-.70	-.37
Self-esteem	1978	.08	.02	.26	.03
BSRI (femininity) #	1983	.39	.28	-.30	-.68
Core tie	1983	-.01	-	.43	-
IBM (care)	1983	-	-.03	-	-.35
IBM (control)	1983	-	.12	-	.09
Group means	non-case	.33	.28	.24	.38
	case	-1.05	-.97	-.80	-1.32
Wilks' Lambda		.742	.783	.836	.657
Chi square		29.29	18.98	8.26	15.77
Significance		<.005	ns	ns	ns
Sensitivity		73.1%	68.4%	69.2%	76.9%
Specificity		79.5%	74.2%	62.8%	83.7%
Correct classification		78.0%	72.9%	64.3%	82.1%

\* p <.05

# BSRI = Bem Sex Role Inventory, femininity subscale  
 Self-esteem: higher scores reflect low self-esteem  
 Core tie : higher score indicates presence of an intimate  
 IBM care : higher score indicates low care  
 IBM control: higher score indicates high control

Table 3 Discriminant Function Analysis using  
RDC cases and noncases at 1983

Variable		Standardised Discriminant Function Coefficients			
	Year	1983 Lifetime Cases		1978-83 New Cases	
Sex		.11	.09	.04	.07
Maternal care	1978	.20	.32	.42	-.57
Paternal care	1978	.52*	.53*	.35	-.26
Maternal protection	1978	-.36	-.10	-.36	.12
Paternal protection	1978	.32	.32	.29	-.39
Neuroticism	1978	-.03	.02	.14	-.20
Trait depression	1978	-.54*	-.48	-.49	.35
Dependency	1978	-.05	.07	-.27	.15
Self-esteem	1978	.07	.07	.08	.00
BSRI (femininity) #	1983	-.25	-.26	.23	.27
Core tie	1983	.17	-	-.17	-
IBM (care)	1983	-	-.17	-	.28
IBM (control)	1983	-	.24	-	.20
Group means	non-case	.38	.37	.22	-.23
	case	-.52	-.55	-.59	.66
Wilks' Lambda		.832	.928	.882	.866
Chi square		27.72	23.01	14.73	14.01
Significance		<.005	<.05	ns	ns
Sensitivity		67.6%	69.2%	65.7%	70.4%
Specificity		72.3%	71.8%	67.0%	69.2%
Correct classification		70.3%	70.8%	66.7%	69.5%

\* p <.05

# BSRI = Bem Sex Role Inventory, femininity subscale  
Self-esteem: higher scores reflect low self-esteem  
Core tie : higher score indicates presence of an intimate  
IBM care : higher score indicates low care  
IBM control: higher score indicates high control

Table 4 Discriminant Function Analysis using  
RDC cases and noncases at 1983

Variable	Year	Standardised Discriminant Function Coefficients			
		Female	Female	Male	Male
Maternal care	1978	.12	.18	-.50	-.32
Paternal care	1978	.56*	.52	-.11	-.29
Maternal protection	1978	-.68*	-.50	-.72	-.66
Paternal protection	1978	.48*	.45	.10	-.34
Neuroticism	1978	-.15	-.11	-.24	-.16
Trait depression	1978	-.40	-.44	.64	.39
Self-esteem	1978	.27	.17	.29	.06
Dependency	1978	-.03	.10	.48	.39
BSRI (femininity) #	1983	-.14	-.13	.53	.53
Core tie	1983	.20	-	-.03	-
IBM (care)	1983	-	-.22	-	.27
IBM (control)	1983	-	.14	-	.64
Group means	non-case	.54	.45	-.26	-.26
	case	-.70	-.68	.40	.41
Wilks' Lambda		.722	.762	.899	.731
Chi square		31.95	21.05	4.85	8.15
Significance		<.001	<.05	ns	ns
Sensitivity		72.9%	64.7%	60.9%	72.2%
Specificity		77.0%	78.4%	60.6%	77.8%
Correct classification		75.2%	72.9%	60.7%	75.6%

\* p <.05

# BSRI = Bem Sex Role Inventory, femininity subscale  
Self-esteem: higher scores reflect low self-esteem  
Core tie : higher score indicates presence of an intimate  
IBM care : higher score indicates low care  
IBM control: higher score indicates high control

Table 5 Discriminant Function Analysis using  
'defined' cases and noncases at 1983

Variable	Standardised Discriminant Function Coefficients				
	Year	1983 Lifetime Cases		1978-83 New Cases	
Sex		.12	.14	.26	.27
Maternal care	1978	-.04	.11	.08	.31
Paternal care	1978	.67*	.71*	.74*	.68*
Maternal protection	1978	-.32	-.18	-.40	-.14
Paternal protection	1978	.24	.21	.42	.29
Neuroticism	1978	-.12	.03	-.13	-.07
Trait depression	1978	-.60*	-.61*	-.52*	-.54*
Dependency	1978	-.06	-.06	-.11	-.09
Self-esteem	1978	.26	.41	.33	.41
BSRI (femininity) #	1983	-.05	.06	-.08	-.14
Core tie	1983	.29	-	.17	-
IBM (care)	1983	-	.06	-	.24
IBM (control)	1983	-	-.27	-	-.21
Group means	non-case	.35	.32	.25	.22
	case	-.80	-.86	-.99	-.93
Wilks' Lambda		.780	.782	.797	.776
Chi square		37.41	20.00	29.55	27.62
Significance		<.001	<.005	<.005	<.01
Sensitivity		75.0%	74.3%	75.0%	77.3%
Specificity		73.5%	73.7%	76.1%	77.9%
Correct classification		73.9%	73.9%	75.9%	77.8%

\* p <.05

# BSRI = Bem Sex Role Inventory, femininity subscale  
Self-esteem: higher scores reflect low self-esteem  
Core tie : higher score indicates presence of an intimate  
IBM care : higher score indicates low care  
IBM control: higher score indicates high control

Table 6 Discriminant Function Analysis using  
'defined' cases and noncases at 1983

Variable	Standardised Discriminant Function Coefficients				
	Year	Female	Female	Male	Male
Maternal care	1978	.05	.22	.20	.17
Paternal care	1978	<b>.70*</b>	<b>.67*</b>	<b>-.36</b>	<b>-.55</b>
Maternal protection	1978	<b>-.53*</b>	<b>-.38</b>	<b>.32</b>	<b>-.44</b>
Paternal protection	1978	<b>.39</b>	<b>.28</b>	<b>-.13</b>	<b>-.51</b>
Neuroticism	1978	.10	.11	.26	.27
Trait depression	1978	<b>-.54*</b>	<b>-.57*</b>	<b>.48</b>	<b>.36</b>
Dependency	1978	<b>-.08</b>	<b>-.09</b>	<b>.42</b>	<b>.38</b>
Self-esteem	1978	<b>.47*</b>	<b>.57*</b>	.20	.05
BSRI (femininity) #	1983	.01	.03	<b>.42</b>	<b>.46</b>
Core tie	1983	.22	-	<b>-.41</b>	-
IBM (care)	1983	-	<b>-.02</b>	-	.06
IBM (control)	1983	-	<b>-.01</b>	-	<b>.63</b>
Group means	non-case	.48	.42	<b>-.27</b>	<b>-.38</b>
	case	<b>-1.06</b>	<b>-1.14</b>	<b>.68</b>	<b>1.06</b>
Wilks' Lambda		.658	.670	.841	.702
Chi square		41.02	31.04	7.67	13.26
Significance		<.001	<.005	ns	ns
Sensitivity		77.0%	78.3%	64.7%	47.1%
Specificity		82.9%	79.0%	66.7%	79.5%
Correct classification		78.9%	78.6%	66.1%	69.6%

\* p <.05

# BSRI = Bem Sex Role Inventory, femininity subscale  
Self-esteem: higher scores reflect low self-esteem  
Core tie : higher score indicates presence of an intimate  
IBM care : higher score indicates low care  
IBM control: higher score indicates high control

Table 7 Discriminant Function Analysis using  
DSM-III lifetime cases and noncases at 1988

Variable		Standardised Discriminant Function Coefficients			
	Year	1988 Lifetime Cases		1978-88 New Cases	
Sex		-.06	-.20	.15	.16
Maternal care	1978	.13	.22	.13	.25
Paternal care	1978	-.45	-.33	-.44	-.10
Maternal protection	1978	.60*	.68*	.61	.56
Paternal protection	1978	-.65*	-.47	-.74*	-.32
Neuroticism	1978	.31	.32	.28	.19
Dependency	1978	.04	-.06	.15	.14
Trait depression	1978	.14	.11	.07	.11
Self-esteem	1983	.41	.41	.32	.52
BSRI (femininity) #	1983	.01	-.11	.19	.05
Core tie	1988	-.38	-	-.51	-
3 or more children	1988	.10	.11	.29	.36
IBM (care)	1988	-	.24	-	.22
IBM (control)	1988	-	.16	-	.29
Group means					
	non-case	-.28	-.22	-.20	-.22
	case	.64	.59	.62	.65
Wilks' Lambda		.845	.884	.890	.911
Chi square		24.65	14.56	15.57	10.13
Significance		<.05	ns	ns	ns
Sensitivity		59.6%	50.0%	61.8%	48.0%
Specificity		75.8%	71.7%	70.1%	66.3%
Correct classification		69.5%	65.9%	68.9%	62.4%

\* p < .05

# BSRI = Bem Sex Role Inventory, femininity subscale  
Self-esteem: higher scores reflect low self-esteem  
Core tie: higher score indicates presence of an intimate  
3 children: higher score indicates presence  
IBM care: higher score indicates low care  
IBM control: higher score indicates high control



Table 8

Discriminant Function Analysis using  
DSM-III lifetime cases and noncases at 1988

Variable		Standardised Discriminant Function Coefficients			
	Year	Female	Female	Male	Male
Maternal care	1978	.18	.11	-.12	-.01
Paternal care	1978	-.42	-.15	-.60	-.85
Maternal protection	1978	.93*	.90*	-.33	-.42
Paternal protection	1978	-.67	-.34	-.88*	-1.01*
Neuroticism	1978	.42	.52	.28	.28
Dependency	1978	-.01	-.17	.49	.38
Trait depression	1978	.05	.00	.07	.58
Self-esteem	1983	.10	.18	.82*	.58
BSRI (femininity) #	1983	-.18	-.33	-.18	.71
Core tie	1988	-.22	-	-.68*	-
3 or more children	1988	-.02	-.08	.48	.52
IBM (care)	1988	-	.06	-	-.04
IBM (control)	1988	-	.23	-	.05
Group means	non-case	-.69	-.39	-.43	-.48
	case	.60	.79	1.12	1.59
Wilks' Lambda		.779	.795	.664	.555
Chi square		24.11	18.16	17.40	18.23
Significance		<.05	ns	ns	ns
Sensitivity		65.7%	62.9%	60.0%	53.3%
Specificity		76.7%	75.3%	84.2%	89.5%
Correct classification		73.1%	71.3%	77.4%	79.3%

\* p &lt; .05

# BSRI = Bem Sex Role Inventory, femininity subscale  
 Self-esteem: higher scores reflect low self-esteem  
 Core tie : higher score indicates presence of an intimate  
 3 children: higher score indicates presence  
 IBM care : higher score indicates low care  
 IBM control: higher score indicates high control

Table 9 Discriminant Function Analysis using  
RDC cases and noncases at 1988

Variable		Standardised Discriminant Function Coefficients			
	Year	1988 Lifetime Cases		1978-88 New Cases	
Sex		-.15	-.26	-.13	-.18
Maternal care	1978	-.05	-.02	-.08	-.02
Paternal care	1978	-.42*	-.43	-.27	-.16
Maternal protection	1978	.42	.41	.44	.39
Paternal protection	1978	-.34	-.46	-.46	-.49
Neuroticism	1978	-.16	.01	-.09	-.05
Trait depression	1978	.30	.44	.32	.36
Dependency	1978	-.06	.09	.21	.24
Self-esteem	1983	.47*	.50*	.48*	.51*
BSRI (femininity) #	1983	.30	.36	.32	.38
Core tie	1988	-.24	-	-.38	-
3 or more children	1988	.21	.25	.34	.37
IBM (care)	1988	-	.15	-	.32
IBM (control)	1988	-	-.03	-	-.03
Group means	non-case	-.56	-.46	-.39	-.38
	case	.56	.53	.67	.71
Wilks' Lambda		.762	.802	.789	.745
Chi square		39.75	25.91	27.03	22.89
Significance		<.001	<.05	<.005	<.05
Sensitivity		61.0%	57.6%	60.0%	58.3%
Specificity		77.9%	76.1%	72.7%	73.1%
Correct classification		69.5%	67.5%	68.0%	68.0%

\* p <.05

# BSRI = Bem Sex Role Inventory, femininity subscale  
Self-esteem: higher scores reflect low self-esteem  
Core tie : higher score indicates presence of an intimate  
3 children: higher score indicates presence  
IBM care : higher score indicates low care  
IBM control: higher score indicates high control

Table 10 Discriminant Function Analysis using  
RDC cases and noncases at 1988

<u>Variable</u>		<u>Standardised Discriminant Function Coefficients</u>			
	Year	<u>Female</u>	<u>Female</u>	<u>Male</u>	<u>Male</u>
Maternal care	1978	.03	.01	-.34	.04
Paternal care	1978	-.55*	-.52	-.16	-.25
Maternal protection	1978	.58*	.58*	.01	-.06
Paternal protection	1978	-.45	-.45	-.65	-.71
Neuroticism	1978	.27	.14	-.11	-.04
Trait depression	1978	.29	.34	.25	.27
Dependency	1978	-.07	-.08	.68	.49
Self-esteem	1983	.26	.36	.95*	.80*
BSRI (femininity) #	1983	.25	.31	.50	.40
Core tie	1988	-.14	-	-.42	-
3 or more children	1988	.20	.20	.21	.54
IBM (care)	1988	-	.12	-	-.17
IBM (control)	1988	-	-.28	-	-.69
Group means	non-case	-.70	-.66	-.47	-.56
	case	.60	.58	.62	.88
Wilks' Lambda		.701	.735	.757	.659
Chi square		34.29	24.36	11.84	12.98
Significance		<.001	<.05	ns	ns
Sensitivity		67.2%	74.1%	72.7%	72.7%
Specificity		80.0%	80.0%	74.2%	80.6%
Correct classification		73.2%	76.9%	73.6%	77.4%

\* p <.05

# BSRI = Bem Sex Role Inventory, femininity subscale  
Self-esteem: higher scores reflect low self-esteem  
Core tie : higher score indicates presence of an intimate  
3 children: higher score indicates presence  
IBM care : higher score indicates low care  
IBM control: higher score indicates high control

Table 11 Discriminant Function Analysis using  
'defined' cases and noncases at 1988

Variable		Standardised Discriminant Function Coefficients			
	Year	1988 Lifetime Cases		1978-88 New Cases	
Sex		.16	.27	.25	.28
Maternal care	1978	.07	.17	.14	.29
Paternal care	1978	.52*	.51*	.61*	.54*
Maternal protection	1978	-.31	-.23	-.37	-.24
Paternal protection	1978	.31	.23	.49*	.37
Neuroticism	1978	-.14	.01	-.15	-.08
Dependency	1978	-.01	.06	-.03	-.03
Trait depression	1978	-.46*	-.54*	-.41	-.43
Self-esteem	1983	-.17	-.09	-.08	-.09
BSRI (femininity) #	1983	.06	.04	-.05	-.14
Core tie	1988	.42*	-	.44*	-
3 or more children	1988	-.14	-.16	-.10	-.15
IBM (care)	1988	-	.25	-	.18
IBM (control)	1988	-	.00	-	.10
Group means	non-case	.39	.31	.29	.24
	case	-.85	-.88	-1.06	-1.10
Wilks' Lambda		.751	.780	.762	.791
Chi square		41.84	29.17	34.61	24.49
Significance		<.001	<.01	<.005	<.05
Sensitivity		66.7%	66.7%	72.4%	80.0%
Specificity		75.5%	74.2%	75.5%	75.3%
Correct classification		72.7%	72.2%	74.8%	76.1%

\* p <.05

# BSRI = Bem Sex Role Inventory, femininity subscale  
Self-esteem: higher scores reflect low self-esteem  
Core tie : higher score indicates presence of an intimate  
3 children: higher score indicates presence  
IBM care : higher score indicates low care  
IBM control: higher score indicates high control

Table 12 Discriminant Function Analysis using  
'defined' cases and noncases at 1988

Variable		Standardised Discriminant Function Coefficients			
	Year	Female	Female	Male	Male
Maternal care	1978	.20	.24	.11	.07
Paternal care	1978	.61*	.52*	-.44	-.27
Maternal protection	1978	-.52*	-.49	.20	.64
Paternal protection	1978	.46*	.30	-.56	-.41
Neuroticism	1978	-.10	.02	.37	.49
Dependency	1978	.05	.19	.52	.45
Trait depression	1978	-.36	-.42	.41	.35
Self-esteem	1983	.20	.15	.80*	.65*
BSRI (femininity) #	1983	.14	.07	.43	.40
Core tie	1988	.35	-	-.70*	-
3 or more children	1988	-.04	-.03	.58*	.86*
IBM (care)	1988	-	-.17	-	.16
IBM (control)	1988	-	.07	-	.35
Group means	non-case	.51	.45	-.53	-.53
	case	-1.05	-1.18	1.37	1.77
Wilks' Lambda		.648	.647	.570	.503
Chi square		41.91	34.44	23.89	21.29
Significance		<.001	<.001	<.05	<.05
Sensitivity		75.0%	75.0%	80.0%	66.7%
Specificity		80.6%	77.8%	81.6%	92.1%
Correct classification		78.7%	76.9%	81.1%	84.9%

\* p <.05

# BSRI = Bem Sex Role Inventory, femininity subscale  
Self-esteem: higher scores reflect low self-esteem  
Core tie: higher score indicates presence of an intimate  
3 children: higher score indicates presence  
IBM care: higher score indicates low care  
IBM control: higher score indicates high control

## Is sex necessarily a risk factor to depression?

KAY WILHELM<sup>1</sup> AND GORDON PARKER

*From the School of Psychiatry, University of New South Wales, and Mood Disorders Unit, Prince Henry Hospital, Sydney, Australia*

**SYNOPSIS** To isolate and quantify possible determinants of any increased prevalence of depressive disorders in women we studied a select group of men and women, initially similar in terms of a number of putative social determinants of depression, and reviewed the sample five years later when social role diversity was anticipated. We used the Diagnostic Interview Schedule (DIS) to generate DSM-III and RDC diagnoses to estimate lifetime depressive disorders, and established (via corroborative reports) the likely accuracy of those data. Despite lifetime depression being a relatively common experience, no significant sex differences in depressive episodes were demonstrated, suggesting the possible irrelevance of biological factors in determining any sex difference. As there was not major social role divergence over the five year study, we interpret the lack of a sex difference as a consequence, and suggest that findings support the view that social factors are of key relevance in determining any female preponderance in depression described in general population studies.

### INTRODUCTION

Over the past decade considerable interest has been generated by reports of an increased prevalence of non-endogenous depressive disorders among women, both in community and clinical samples (Weissman & Klerman, 1977; Boyd & Weissman, 1981; Briscoe, 1982; Weissman *et al.* 1984; Jenkins, 1985; Jorm, 1987). The review by Weissman & Klerman considered a number of explanations, broadly divided into artefactual (e.g. over-reporting by women) and real factors, before considering biological, genetic and social determinants of any real difference.

As exceptions to the female preponderance have been described in isolated studies in developing countries (Weissman & Klerman, 1977), in college or university students (Parker, 1979; Hammen & Padesky, 1977) and in a young, employed group (Jenkins, 1985) of men and women matched for age, education and occupation, socio-cultural explanations have been advanced. A quantitative analysis (Jorm, 1987) of the published literature established that there is little sex difference in depression in either

childhood or advanced old age, so that the sex difference is greatest during the period when male and female occupational and social roles diverge the most, again favouring a social role explanation. Rising rates of depression in men, with a consequent tendency towards more equal rates of depression between the sexes, have been shown in four longitudinal studies (Murphy, 1986), and have also promoted interest in social determinants. As most of the studies offering a social role explanation have focused on depressive symptoms, rather than depressed 'cases', the possible relevance of social factors to the sex difference in depressive disorders appears a priority issue.

We judge that the appropriate test of the social determinant hypothesis is to examine for any variation in depressive disorders between the sexes when the putative social determinants are firstly controlled and then non-controlled, and now report such a study. We elected to study a group of men and women, initially homogeneous in terms of age, marital state, social class and occupation, and subsequently after social role diversity was anticipated. We selected a group of teacher trainees, as a previous study looking at the preceding year's students engaged in a similar course at this college (Parker, 1979) had suggested similar levels of depression for the sexes during that university

<sup>1</sup> Address for correspondence: Dr Kay Wilhelm, School of Psychiatry, Prince Henry Hospital, Little Bay 2036, Sydney, Australia

year when social roles were somewhat similar. In our present study, we anticipated no sex difference at the initial assessment, but hypothesized that differences would subsequently emerge in proportion to subsequent social role divergence.

A predictable concern was to measure episodes of depression in a rigorous and clinically meaningful way, no easy task when the issue of what is a 'case' bedevils psychiatric epidemiology and research. Additionally, when we commenced intake into our study in 1978, nosological systems (e.g. ICD-9, DSM-II) used very general and non-specific descriptive statements to generate diagnoses. Being aware in 1978 that case-finding lifetime prevalence measures were being developed, we decided to delay such judgements to a follow-up review and to assess our subjects at intake only on a number of depression inventories and general interview variables.

Subsequently, we chose to use the Diagnostic Interview Schedule (DIS) (Robins *et al.* 1981) as our case-finding technique at a 5-year review. Use of the DIS had major advantages in that it can generate both Research Diagnostic Criteria (RDC) (Spitzer *et al.* 1981) and DSM-III (APA, 1980) point, period and lifetime prevalence diagnoses of major and minor depressive disorders, so allowing comparison with other epidemiological studies, including the Epidemiologic Catchment Area (ECA) study (Regier *et al.* 1984) where it was the case-finding instrument. We elected to have one psychiatrist (K.W.) administer the DIS to all subjects, so as to control for diagnostic and observation variance generated by multiple raters. The DIS had not been validated as a case-finding measure before commencing the ECA study and remains unvalidated at present (Parker, 1987). Therefore, we decided to obtain corroborative information to establish the extent to which DIS-derived data on occurrence and impact of depressive disorders corresponded with judgments of witnesses.

We anticipated that depressive disorders would be less severe in our sample (compared to depression in a clinical group) and that many would cluster on the boundary between major and minor depressive disorders, and between 'caseness' and 'non-caseness'. The RDC system has several advantages over DSM-III in non-

clinical groups in this regard, in having wider boundaries between depressive categories (minor depression requiring two symptoms, as against probable major and definite major depression requiring four and five symptoms respectively). Additionally, by imposing impaired function and help-seeking criteria, it aids a focus on more 'clinical' expressions of depression. Apart from an implicit assumption of impairment in DSM-III (1980) categories (p. 6), the DSM-III depressive disorders do not (apart from adjustment disorder) include specific impairment criteria, because DSM-III was designed for use by psychiatrists in clinical settings where help-seeking and impairment criteria would, by definition, be met and therefore be redundant. Such a difference suggests, however, that DSM-III 'caseness' in general population samples may be less clinically relevant than 'cases' defined by the RDC system. In recognition of these differences, we elected to generate both DSM-III and RDC diagnoses and, to assist consideration of 'caseness' definitions for each system, we have tabulated their criteria and two minor modifications imposed by us in Table 1. With the use of the DIS, one can generate 'significant cases' of DSM-III disorders by imposing the RDC impairment criteria. In this paper, DSM-III rates are reported without the imposition of these help-seeking/impairment criteria, as RDC rates are being simultaneously reported, and it allowed an opportunity to examine sex differences if help-seeking was in fact a determining factor.

## SUBJECTS AND METHODS

### Sample selection

In September 1978, 380 students who had completed a basic Arts or Science university course, and who were then undertaking a one-year teachers' training programme, were approached in class and invited to participate in a five-year research project. While it was explained that information on their depressive experience would be sought over time, they were not informed of the research workers' focus on sex differences in depressive experience, or of the study of a similar group of trainees in the preceding year (Parker, 1979).

Three hundred and fifty of these students successfully completed a questionnaire which

Table 1. RDC and DSM-III criteria for depressive disorders with study differences noted

	Minimal duration	Number of symptoms required	Impaired functioning and help-seeking criteria required
<b>RDC diagnoses</b>			
Major depression:			
probable	1-2 weeks*	4/8	Yes
definite	2 weeks	5/8	Yes
Minor depression:			
definite	2 weeks	2/16†	Yes
Intermittent depression	At least hours and present for at least 2 years	2/16†	Yes
<b>DSM-III diagnoses</b>			
Major depression:			
total	2 weeks	4/8	No
significant	2 weeks	4/8	No‡
Adjustment disorder	Not specified*	Not specified	No‡
Dysthymia	2 years	3/13†	No‡

\* A minimum of two weeks imposed.

† We limited the symptoms to eight, being those listed for RDC major depression.

‡ RDC help-seeking/impairment criteria imposed with use of DIS

included a self-esteem scale (Rosenberg, 1965), assessment of paternal occupation to rate social class on a four-point rating scale (Congalton, 1969), and a question assessing whether they were willing to take part in the longitudinal project. Of the 197 students expressing provisional acceptance, 170 completed and returned a mailed baseline questionnaire and were regarded as the study cohort. As a brief screen to determine if the 170 study subjects (114 female, 56 male) differed from the remaining 181 students (119 female, 62 male) who had finally elected not to take part, scores on the self-esteem scale were compared. The mean scores for the respective groups were 1.53 and 1.43, the differences not being significant ( $t = 0.74$ , NS). Male 'refusers' did not differ from male sample members on the self-esteem measure ( $t = 1.39$ ) or in age ( $t = 1.32$ ). Similarly, female 'refusers' did not differ from female sample members on the self-esteem measure ( $t = 0.04$ ) or in age ( $t = 0.46$ ). Males and females did not differ in their likelihood of either refusing or joining the sample ( $\chi^2 = 0.06$ ).

#### Entry assessment

At entry, we compared male and female members of the cohort on a number of measures assessing depressive experience, including state (Wilson, 1979) and trait (Costello & Comrey, 1967) depression scales, the latter, according to its

designers, being a measure assessing a 'person's tendency to experience a depressive mood'. Finally, we gave two personality measures, a dependency scale derived from the Depressive Experiences Questionnaire (Blatt *et al.* 1975) and the Eysenck Personality Inventory neuroticism scale (Eysenck & Eysenck, 1964).

#### Five-year review

In November 1983, five years after intake, a concerted effort was made to obtain follow-up data on all 170 subjects in the cohort, and 165 (97.1%) were successfully located. From December 1983 to May 1984 the first author was able to interview 150 (91%) of these while, for the remaining 15 (9.0%), data were collected from mailed questionnaires. For most of those in the latter group who were no longer in Sydney, telephone contact was made to clarify details. To evaluate the acceptability and accuracy of such non-interview derived data, 10 subjects (three female, seven male) were requested to complete a mailed version of the DIS assessing anxiety and depression symptoms, and were subsequently interviewed in person with the first author administering the DIS. For nine of the 10 subjects the judgements made at interview were the same as those made from the mail-generated data, although there was a tendency for subjects to report more symptoms at interview. For one subject, an episode of major depression was



elicited at interview that had not been otherwise reported. Overall, we concluded that collection of data by mail was likely to lead to a slight underrepresentation of episodes of depression in those who were possible 'cases'. Nevertheless, in order to achieve a high response rate for those entering the study 5 years earlier, and because any bias was small and affecting only 15 of the sample, we elected to include their data but decided, for those reporting any depressive symptoms in their mailed questionnaires, to seek careful clarification by telephone contact.

None of the 165 subjects located refused to be interviewed, although, after interview (including the DIS), one subject decided not to complete the self-report questionnaires and not to have any further involvement, stating that material discussed in the interview had reminded her of previous depressive episodes that she wished to forget.

The five-year follow-up data were collected by the first author, a psychiatrist then with 12 years of clinical experience, and trained in administration of the DIS. The follow-up comprised completion of a number of self-report measures, and a comprehensive semi-structured interview which established a framework for an informal psychiatric, family and developmental history to be taken, before administration of those sections of the DIS dealing with anxiety states and depression.

We need now to make some specific comments about our use of the DIS (See Table 1). While DSM-III uses diagnostic criteria for major depression and dysthymia, it merely offers a brief description in defining a minor depressive disorder such as 'adjustment disorder with depressed mood', so allowing considerable subjectivity in rating. While the DIS is not designed to generate diagnoses for minor depressive disorders, the standardized questions generate material readily encapsulated by the RDC definition of minor depression. We elected to allow a 'case' of RDC minor depression if there had been a depressive episode lasting at least 2 weeks and possessing two of the eight DSM-III symptoms used for major depression with severity criteria also fulfilled. This allowed for inclusion of subjects who reported a depressive disorder with 2-3 of the eight allowed symptoms but a duration of less than 2 years. In such instances, a diagnosis of RDC definite minor

depression was used, but categorization was more problematical for a DSM-III diagnosis which requires that the reaction be 'maladaptive' to qualify for a diagnosis of adjustment disorder, allowing interviewer variance in assessing maladaptation. Additionally, for adjustment disorder, there is no minimum number of symptoms stated, so, for our purposes, we required that the subject judge the episode as significant, with a duration of at least two weeks, precipitated within three months of a psychosocial stressor and fulfilling the RDC impairment criteria.

At interview, each subject was asked to nominate another person who had known the subject well for a number of years and who might be able to provide an independent account of the subject's depressive experience. Consent and a completed corroborative interview were achieved for 133 of the 150 (88.7%) subjects interviewed in person, and therefore 80.6% of the whole cohort. Where and when possible, the nominated witness was interviewed immediately to avoid discussion and 'priming' by the subjects, and this occurred for 56 of the 133 corroborative interviews. If the person nominated was not available, the subject was asked to inform that witness that there would be telephone contact in the next few days but not to describe the exact nature of the information to be sought, this procedure being adopted for the remaining 77 subjects. On a number of occasions more than one witness was sought to clarify details.

Whether contact was in person or by phone, a similar approach was used. The objectives of the research were briefly explained (but without reference to the issue of sex differences) and the following question was put: 'Has X, in the time that you have known him/her, ever experienced an episode of depression lasting at least two weeks, when he/she seemed depressed or sad or behaved very differently from normal or gave you cause for concern?'. If the answer was negative, the informant was prompted once more 'are you sure that.....' and if again negative, no further exploration occurred. If the answer was positive, the informant was then asked: 'Could you tell me when the episode(s) occurred and something about it (them)'. Information was sought as to whether there was a persistent and qualitative difference in mood and behaviour, the timing of onset and duration

of episodes, and the impact of the episode on the subject and others. If the informant volunteered details of possible causes for such episodes, this was noted but not otherwise pursued. Information concerning timing and severity of episodes was compared by the interviewer at that time to that given by the subject, without the informant being aware of what the subject had reported. If there were discrepancies (e.g. an episode being noted in a different year), further questions were asked for clarification.

A number of questionnaires completed in 1978 (assessing self-esteem, state depression, trait depression, dependency, and neuroticism) were readministered, together with two additional measures considered here, one assessing occupational satisfaction and importance (Renick & Lawler, 1978). The other was the BEM (Bem, 1974) sex role inventory designed to rate masculinity and femininity as two independent dimensions, so that subjects may be characterized as masculine, feminine, androgynous or neutral. This measure was included to examine the relevance of sex roles in our sample.

## RESULTS

We report data on the 165 subjects assessed at entry and five-year review, comprising 109 females and 56 males. The mean age of the cohort was 23.4 years at entry, and there was no difference in mean ages of the males and females. A total of 22 subjects acknowledged then that they had previously consulted a primary phys-

ician or a psychiatrist for an episode of depression, the consultation rate being similar ( $\chi^2 = 0.29$ , NS) for females (14.6%) and males (10.7%), although women were more likely to have sought help from a friend for depression ( $\chi^2 = 5.81$ ,  $P < 0.05$ ). Medication for 'nerves' had been taken by 11.9% of the females and 7.1% of the males ( $\chi^2 = 0.55$ , NS). Table 2 shows that mean scores on the state depression, trait depression and self-esteem scales were quite similar for males and females, both in 1978 and in 1983. Females scored significantly higher than males on the dependency and neuroticism measures, and on both occasions of testing. The only change over time on any of the tabulated measures was for self-esteem to improve significantly, and for both sexes (females:  $t = 3.37$ ; males;  $t = 3.49$ ; both  $P < 0.001$ ). The occupational scale established that the females rated their principal work as both more satisfying and as more important than did the males in 1983. While there were trends for the females to score more highly than males on the BEM femininity sex role scale and for males to score more highly on the masculinity scale, neither of these trends was significant.

### Sociodemographic data at 1978 and 1983

Table 3 reports socio-demographic data. At baseline, the sexes did not differ on any of the socio-demographic variables examined (e.g. marital status, number of children, employment, age and paternal social class), although, if subjects are dichotomized (classes 1 and 2 v.

Table 2. Mean scores on several depression and personality scales

Measure	Assessment	Female	Male	<i>t</i> test	Significance
Self-esteem*	1978	1.6	1.5	0.63	NS
	1983	1.1	0.7	1.73	NS
Trait depression	1978	30.5	32.3	0.90	NS
	1983	30.9	31.7	0.44	NS
State depression	1978	57.2	56.7	0.49	NS
	1983	55.5	54.6	0.92	NS
Dependency	1978	53.9	50.4	2.24	< 0.05
	1983	54.2	50.1	2.89	< 0.01
Neuroticism	1978	9.5	8.0	1.97	< 0.05
	1983	9.5	7.3	2.57	< 0.025
Dysfunctional attitudes	1983	78.4	78.4	0.00	NS
BEM Scale					
Masculinity	1983	4.54	4.74	1.85	NS
Femininity	1983	4.83	4.65	2.16	NS
Occupational satisfaction	1983	89.1	81.8	2.79	< 0.01
Occupational importance	1983	95.0	89.4	2.60	< 0.01

\* Higher scores reflect a lower self-esteem.

Table 3. *Socio-demographic characteristics of the sample, by sex*

Variable	Year of assessment	Female N = 109	Male N = 56	Sex difference
Marital state				
Married	1978	23	9	$\chi^2 = 0.31$
	1983	58	25	$\chi^2 = 0.76$
Never married	1978	80	45	$\chi^2 = 0.64$
	1983	32	24	$\chi^2 = 2.43$
Other (e.g. divorced, de facto, widowed)	1978	6	2	$\chi^2 = 0.03$
	1983	19	7	$\chi^2 = 0.36$
Partner status				
Living with partner	1978	23	11	$\chi^2 = 0$
	1983	66	30	$\chi^2 = 0.48$
Age	1978	23.1	23.9	$t = -1.21$
	1983	29.1	30.1	$t = -1.51$
Children				
Number with children	1978	9	4	$\chi^2 = 0$
	1983	39	17	$\chi^2 = 0.27$
Mean number of children	1978	0.20	0.10	$t = 0.53$
	1983	0.56	0.57	$t = -0.07$
Employment				
Full-time work (81% teaching)	1979*	83	47	$\chi^2 = 0.92$
Full-time work (78% teaching)	1983	87	50	$\chi^2 = 1.72$
Part-time work (83% teaching)	1979*	21	8	$\chi^2 = 0.34$
Part-time work (65% teaching)	1983	13	4	$\chi^2 = 0.48$
Unemployed	1979	12	3	$\chi^2 = 0.83$
	1983	13	6	$\chi^2 = 0$
Home duties	1979	2	0	$\chi^2 = 0.07$
	1983	11	0	$\chi^2 = 4.54^\dagger$
Experience of unemployment during study		29	17	$\chi^2 = 0.11$
Mean duration of unemployment in weeks		24.4	21.3	$t = 0.46$
Social class	1978	1	4	
	2	46	19	$\chi^2 = 4.98$
	3	39	29	
	4	6	2	

\* First year of work.

†  $P < 0.05$ .

classes 3 and 4) there is a trend for the males to be slightly lower on social class ( $\chi^2 = 3.38$ , df 1, NS) as estimated from paternal occupation. At the five-year review, 75% were in the 26–29 year age range, with only four being forty years or older. Fifty-three per cent of the females and 45% of the males were then married as against 21% and 16% respectively at the intake assessment, with the likelihood of having married over the interval being similar for the two sexes. Similarly, while less than 10% were parents in 1978 as against a third in 1983, the increased parenthood rate was similar for males and females ( $\chi^2 = 0$ ). Neither the full-time employment rate in 1983 ( $\chi^2 = 0$ ) nor the unemployment rate over the interval ( $\chi^2 = 0.61$ ) showed any sex difference. There was a differential in being engaged in home duties, however,

with no male being so assigned over the five years but with an increase from two to eleven females so engaged between 1979 and 1983.

By the five-year review, 45 (27% of the group) had at some time of their life sought professional help for psychological problems, although females (28%) were no more likely ( $\chi^2 = 0.27$ , NS) to do so than males (27%). Twenty-eight (17%) had sought such assistance from a primary care physician, 21 (13%) from a social worker or counsellor, and 14 (8.5%) from a psychiatrist.

At interview, four (2.4%) subjects had a current definite or probable major depression and five (3.0%) a minor depression, using RDC diagnoses, with the six-month prevalence for these diagnoses being 3.0% and 4.2% respectively. The DSM-III point and six-month preva-

Table 4. Prevalence and incidence data using RDC diagnoses, by sex

Diagnosis and time interval	Female (N = 109)	Male (N = 56)	Total (N = 165)	$\chi^2$
Cases before intake in 1978				
Major depression (MDD):				
definite	7 (6.4%)	2 (3.6%)	9 (5.5%)	0.16
definite and probable	9 (8.3%)	2 (3.6%)	11 (6.9%)	0.66
Intermittent minor depression (IMD)	3 (2.8%)	0 (0.0%)	3 (1.8%)	0.41
MMD (definite and probable) and IMD	12 (11.0%)	2 (3.6%)	14 (8.5%)	1.77
Minor Depression	16 (14.7%)	10 (17.9%)	26 (15.8%)	0.10
MDD, IMD and minor depression	25 (22.9%)	11 (19.6%)	36 (21.8%)	0.08
New cases - 1978-83				
Major depression (MDD):				
definite	9 (8.3%)	6 (10.7%)	15 (9.1%)	0.05
definite and probable	12 (11.0%)	9 (16.1%)	21 (12.7%)	0.46
Intermittent minor depression (IMD)	3 (2.8%)	3 (5.4%)	6 (3.6%)	0.17
MDD and IMD	15 (13.8%)	11 (19.6%)	26 (15.8%)	0.57
Minor depression	16 (14.7%)	6 (10.7%)	22 (13.3%)	0.22
MDD, IMD and minor depression	23 (21.1%)	12 (21.4%)	35 (21.2%)	0.02
Cases - Total lifetime				
Major depression (MDD):				
definite	16 (14.7%)	8 (14.3%)	24 (14.5%)	0.03
definite and probable	21 (19.3%)	11 (19.6%)	32 (19.4%)	0.02
Intermittent minor depression	6 (5.5%)	3 (5.4%)	9 (5.5%)	0.10
MDD + IMD	27 (24.8%)	13 (23.2%)	40 (24.2%)	0.01
Minor depression	32 (29.4%)	16 (28.6%)	48 (29.1%)	0
Major + Intermittent + Minor	48 (44.0%)	23 (41.1%)	71 (43.0%)	0.04

Table 5. Prevalence and incidence data using DSM-III diagnoses, by sex

	Female (N = 109)	Male (N = 56)	Total (N = 165)	$\chi^2$
Cases before intake in 1978				
Major depression (total)*	10 (9.2%)	2 (3.6%)	12 (7.3%)	1.00
Dysthymia	3 (2.8%)	0 (0.0)	3 (1.8%)	0.41
Major + Dysthymia	13 (11.9%)	2 (3.6%)	15 (9.1%)	2.20
Adjustment disorder	26 (23.9%)	11 (19.6%)	37 (22.4%)	0.17
Major + Dysthymia + Adjustment	25 (22.9%)	10 (17.9%)	35 (21.2%)	0.31
New cases - 1978-83				
Major depression (total)*	17 (15.6%)	9 (16.1%)	26 (15.8%)	0.02
Dysthymia	3 (2.8%)	3 (5.4%)	6 (3.6%)	0.17
Major + Dysthymia	19 (17.4%)	12 (21.4%)	31 (18.8%)	0.17
Adjustment disorder	17 (15.6%)	6 (10.7%)	23 (13.9%)	0.38
Major + Dysthymia + adjustment	28 (25.7%)	13 (23.2%)	41 (24.8%)	0.02
Cases - Total lifetime				
Major depression (total)*	27 (24.8%)	11 (19.6%)	38 (23.0%)	0.29
Dysthymia	6 (5.5%)	3 (5.4%)	9 (5.5%)	0.10
Major depression and dysthymia	32 (29.4%)	14 (25.0%)	46 (27.9%)	0.17
Adjustment disorder	33 (30.3%)	16 (28.6%)	49 (29.7%)	0.01
Major + Dysthymia + Adjustment	54 (49.5%)	24 (42.9%)	78 (47.3%)	0.42

\* Note, rates for major depression (total), are given without imposed impairment criteria, refer to Table 1.

Rates for major depression (significant) are identical with RDC major depression (definite and probable) and are found on Table 4.

ince data were similar, with the low base rate suggesting that analyses of sex differences in rates would be unwise.

Tables 4 & 5 provide pre-intake, total lifetime

and 5-year incidence depression data (all being derived from the data obtained at the review assessment) using the RDC and DSM-III classification systems respectively. In calculating rates

we allocated subjects to more than one single category when relevant but, for combined categories, individuals diagnosed positively on more than one diagnostic category were logically only counted once.

#### Life-time prevalence of depression

The RDC 'definite' case rate of 14.5% and the 'definite and probable' case rate (19.4%) for major depression are somewhat less than that for DSM-III (23.0%) because such a DSM-III diagnosis requires fewer symptoms than RDC 'definite' and imposes no help-seeking or social impairment criteria as in RDC 'probable' cases. For neither the DSM-III or RDC system is there a sex difference for lifetime major or minor depression, nor is there any trend for a sex difference. When all categories of depression are combined to create 'cases' and 'non-cases', we find that almost half of the sample (using either RDC or DSM-III systems) had had a lifetime depressive episode, with the RDC and DSM-III estimates corresponding closely, and with any trend for females to be more likely to report a lifetime depressive episode being non-significant. Subsidiary examination suggested that this trend emerged because women positive for depression more often tended to report either major or minor depressive episodes, whereas men positive for depression more often tended to report both major and minor episodes of depression. In the five-year interval between assessments when the sample formally took up adult responsibilities, there is no suggestion of any female preponderance in onset cases of major depression or combined depressive categories.

The mean length of major depressive episodes was also examined as there is the possibility that women have longer episodes, so accounting for a higher prevalence. While there was a trend for women to experience longer RDC episodes, the mean length being 29.4 weeks (s.d. 33.1) for women and 25.3 weeks (s.d. 20.2) for men, the difference was not significant ( $t = -0.65$ ). For DSM-III diagnoses of major depression, the overall mean length of episodes was shorter, reflecting the inclusion of subjects without impairment criteria. For women, the mean length of reported DSM-III episodes was 26.4 weeks (s.d. 29.0) and for men 22.3 weeks (s.d. 19.7), the difference again being non-significant ( $t = 0.81$ ).

In terms of anxiety disorders, the DIS-generated diagnoses established that nine females and three males had a simple phobia, four females and one male a social phobia, two females and three males agoraphobia, and four females and two males a generalized anxiety disorder, with a number being represented in more than one diagnostic group, so that a diagnosis of anxiety disorder was made for 16 (15%) of the females and six (11%) of the males. Seventeen (77%) of those with an anxiety disorder developed depression concurrent with (12 subjects) or subsequent to (five subjects) their anxiety disorder. It is important to note that no subject received a diagnosis of bipolar disorder.

#### Occurrence of depressive episodes pre 1978, and from 1978 to 1983

In Tables 4 & 5 we report data on new cases of RDC and DSM-III depression before intake assessment in 1978 and over the 5 year interval. While there are non-significant and slight trends for the females to be more likely to have reported a major depressive episode, the rarity of such an episode for each sex suggests caution in any interpretation. Any trend for a higher pre-intake RDC major depression rate for females is countered by a higher trend for males to develop episodes during the study, so that the lifetime rates are very similar.

#### Corroborative reports by nominated witnesses

We examined our subject-generated data against reports by corroborative witnesses, given that we had such a witness for 81% of the cohort. For RDC major depression, we established that 98% of witnesses agreed with the subjects' judgement of it being present or absent, the Kappa coefficient being 0.93. Of the 32 subjects reporting an episode, 26 volunteered contactable witnesses, with 23 of the 26 agreeing with the subject's report in regard to occurrence and impact (impairment). In the three instances of dissonant views, two informants confirmed the timing but considered the episode less disabling than judged by the subject, while one failed to confirm severity or timing.

For the 54 subjects reporting either RDC intermittent minor or minor depression, 49 witnesses were contacted and 41 agreed with the subject's view about timing and impact, the

Kappa coefficient being 0.86. Specifically, for the whole group of 'cases' and 'non-cases', three witnesses judged episodes as having occurred which had not been volunteered by subjects while, for five subjects volunteering episodes, two witnesses had not noted the episode and an additional three agreed with the timing but did not judge the episode as having had any significant impact.

## DISCUSSION

Our study design required selection of a group whose male and female members had a number of important social risk factors to depression initially controlled. As a consequence and because the selected teacher trainees were predominantly middle class, their general experience of depression may differ from that of a more broadly-based general population sample. The actual prevalence estimates for the whole sample are therefore not of central importance. Nevertheless, we need to examine the likely accuracy of our data and their comparability with other prevalence estimates before addressing our main objective, the issue of sex differences.

For those who entered the study, we achieved a very high compliance (97%) with the five-year review and, as the subjects were volunteers, we assume their high motivation favoured the collection of accurate data. As poor agreement between lay interviewers and physician interviewers in DIS diagnoses of major depression has been shown in a number of ECA sub-samples (Folstein *et al.* 1985; Anthony *et al.* 1985; Telzer *et al.* 1985), with lay interviewers 'overdiagnosing' in the first two studies and 'underdiagnosing' in the latter study, we need to consider the likely accuracy of our DIS-generated data. Use of a single interviewer, a trained psychiatrist, for our sample members argues for minimal rater variance, while we believe that the accuracy of DIS data was improved by incorporation of the schedule within a general psychiatric interview.

As the DIS has not been validated as a measure of depressive disorders, we judged that corroborative information should be sought where possible. Such a procedure corresponds somewhat with that adopted by Leckman and colleagues (1982) for the lifetime version of the schedule for Affective Disorders (SADS) (Endi-

cott & Spitzer, 1978), in that they interviewed probands and normals, together with a percentage of their spouses and first degree relatives for a corroborative report, and (for the patients) they consulted medical records to derive a 'best estimate' diagnosis. It is useful to report our experience with corroborative 'witnesses'. Most of our subjects were willing to nominate a witness to be interviewed and, while most of the witnesses had reasonable concerns about privacy, these were allayed with explanation. We were impressed that both parties generally understood there to be differences between sadness and depression, and between functional and dysfunctional depressive states. Corroborative report data were most encouraging, and particularly for major depression, when no witness reported major depression when it was not reported by a subject and only three witnesses failed to report an episode acknowledged by subjects.

Such analyses encourage the view that we obtained accurate estimates of depressive disorders. While offering support for the utility of the DIS, the study cannot however be regarded as having provided a precise test of its validity. The DIS was administered after a semi-structured interview which, at times, addressed depressive experiences and highlighted episodes of depression, so that more information was obtained to generate diagnoses than might have been elicited by the DIS alone. Again, while we interviewed witnesses to establish existence, impact and timing of any episodes, we did not review individual DIS items. In terms of the utility of the DIS, our rater's view is that, if administered by a clinician, the DIS has distinct clarity, is 'user friendly', and assists generation of clinical diagnoses by its firm operational criteria. As noted earlier, significant disparities have been demonstrated in ECA sub-samples between clinician and lay interviewers on the DIS, generally suggesting higher 'case' rates with clinicians. An experienced clinician is perhaps more likely, in comparison to a lay interviewer, to note and pursue non-verbal clues, clarify responses that initially have a 'social desirability' bias, and phrase questions (even standardized ones) with a distinct clinical emphasis on distinguishing functional and dysfunctional features. As a consequence, we believe that clinician raters are more likely to generate

accurate estimates, particularly if the DIS is complemented by open-ended interview and corroborative reports.

There are three important studies of lifetime prevalence worth noting for comparability of data.

In a New Haven study (Weissman & Myers, 1978), 25.8% of the females and 12.3% of males had RDC probable or definite major depression, 11.7% of the females and 5.9% of the males had RDC minor depression, and 33.7% of the females and 17.4% of the males had one or both. Reich *et al.* (1980) determined a lifetime prevalence of major depression as 20–26% for females and 8–12% for men. In the ECA study, Robins *et al.* (1984) calculated a 7.1% rate for females and 3.1% for males in lifetime major depression (being 5.5% for the whole sample and 8.7% in the sub-group aged 25–44 years). Additionally, they calculated a lifetime dysthymia diagnosis for 3.9% of the females and 2.0% of the males (3.0% overall, but 3.8% in the 15–44 year sub-group). Thus, our lifetime rates (RDC female = 44.0%, male = 41.1%; DSM-III female = 49.5%, male = 42.9%), are considerably higher and principally contributed to by the minor categories.

While variation in prevalence estimates obviously depends on the decision rules used by differing categorical systems, our combined male and female major depression prevalence data are generally comparable with previous studies (but not with the ECA rates), while our minor depression rates are considerably higher. Our key finding, against expectation, is the absence of any significant sex difference in lifetime and extended period prevalence data, a finding consistent with the sample's professional utilization rates (assessed both in 1978 and 1983) and supported by corroborative witness data. Equally importantly, we established virtually identical incidence rates for the males and females for depressive disorders over the five years of the study when social role divergence was expected.

Some possible explanations for the similarity of the depression rates for the men and women will be examined. As Jorm (1987) has suggested that the sex difference in depressive states does not peak until the early thirties, it may be that our subjects had not reached the vulnerable age range, although we judge this explanation as

unlikely when Jorm's quantitative synthesis showed a difference emerging in adolescence and being pronounced in the twenties. We intend, however, to continue to review the group at regular intervals, both to examine for any age effect and any effect of differential risk factors to the sexes (a ten-year follow-up is now in progress). As our lifetime depression rates for males were higher than expected from other studies (while the female rates were generally similar), males entering the teaching profession may be a vulnerable group as against the females being a more resilient group, or our data could reflect a recent, and more universal trend for an increasing prevalence of depression in males aged 20–40 years (Murphy, 1986; Hagnell *et al.* 1982). We favour most the possibility that, as social variables effectively remained controlled throughout the study, in that age, marital status, parenthood, and employment (apart from home duties) did not differentiate the sexes, a sex difference was prevented from emerging. We now explore that proposition further by drawing attention to a relevant and rigorous study.

Jenkins (1985) sought to 'assess the likely magnitude of the contribution of biological factors to the reported sex difference in minor psychiatric morbidity by controlling and minimizing' environmental differences. Public servants (university graduates in their late twenties) in the British Home Office were selected. She too chose to examine a group of males and females closely comparable in age, marital status, educational attainments, paternal social class, domestic responsibilities and overall social supports and stresses, but did not formally assess depressive episodes.

The men and women did not differ in terms of the prevalence of minor psychiatric morbidity assessed by the Clinical Interview Schedule (Goldberg *et al.* 1970), but the women recorded slightly higher scores than the men and, on a symptom profile, women were more likely to report depressive symptoms than men (39% v. 28%). Such findings, together with results noted in two studies of young adults (Oliver & Simmons, 1985; Angst & Dobler-Mikola, 1984), suggest that women are more likely to rate as 'cases' if the criterion is number of symptoms and/or if rated symptoms are weighted to more feminine depressive responses such as crying (Parker, 1979) so that any female preponderance

in a sample is likely to be artefactual to some extent (Weissman & Klerman, 1977). If such speculations are valid, then criteria-based diagnostic and classificatory systems may require some modification, as suggested by Angst and Dobler-Mikola (1984) to convert the current weighting given to the number of symptoms in determining 'caseness'. A sex difference in depression might then be expected to emerge more on the basis of any 'real' determinant.

We did, nevertheless, establish some sex differences in our sample: higher scores (both at entry and follow-up) were reported by the women on neuroticism and dependency measures, while scores on the occupational scales suggested that the women found their work more satisfying and important than the men. Neuroticism, as measured by the Eysenck Personality Inventory (Eysenck & Eysenck, 1964) and by related measures, has been held (Katz & McGuffin, 1987) to be a key personality risk factor to depression as it 'encompasses vulnerability to breakdown under stress and a proneness to anxiety and emotional instability'. A personality trait of dependency (Hirschfeld *et al.* 1977) has also been held to dispose to depression. Despite our female subjects scoring higher than the males on these two measures, lifetime depression rates were similar for the two sexes. This finding suggests that if dependency and neuroticism are risk factors, they do not, by themselves, dispose to depression but require interaction with other disposing (e.g. low social class) or triggering (e.g. life events) factors to precipitate a depressive episode.

We failed to demonstrate any differences between the men and the women on the BEM sex role inventory given in 1983, suggesting that sex-typed standards were not a characteristic of this group, be any such similarity an antecedent to, or a consequence of choosing teaching as a career. As this lack of difference intrigues us, we would encourage the use of sex role inventory scales in samples demonstrating sex differences in depressive disorder as such data might provide useful information about mediating mechanisms.

While our generally negative findings suggest that it is unlikely that biological factors have any direct effect on determining sex differences in depression, we must concede and consider several anticipated caveats about this study.

First, there is the possibility of a type II error. There is sufficient power in the sample size to detect differences in the continuous variables where there is an effect size of greater than 0.45 standard deviations and to determine differences in the combined categories of depression if the sex differences were of the order that we had expected (that is a 2:1 female to male sex ratio). There is not sufficient power to reliably detect differences if the individual diagnostic categories are used (Cohen, 1977).

Secondly, might our responders have introduced a bias? Our study design involved us addressing classes of trainee teachers, asking all to complete a brief screening measure anonymously, and noting whether they would be prepared to take part in a longitudinal study of adult development examining depressive disorder together with a number of other issues and, if so, providing their name and address. We would speculate that responders and refusers would differ in terms of basic interest in such a study, and in being comfortable or diffident about the potential intrusiveness. As the stated objective to the classes was to observe a normal group developmentally, with depression being only one of a number of issues mentioned and with no emphasis on psychiatric morbidity, we doubt the possibility of a strong bias influencing acceptance. In support of this speculation we note a recent empirical study by Romans-Clarkson *et al.* (1988), which established that refusers in a community survey of psychiatric disorder did not differ from participants in terms of hospitalized medical or psychiatric illness or in rates of attendance at psychiatric out-patient clinics. Additionally, we did establish that our participants and refusers did not differ on a self-esteem measure, and it might be imagined that such a measure would have differentiated the groups if they differed distinctly in lifetime depressive experience.

Thirdly, are teachers themselves an idiosyncratic group? The answer is almost certainly in the affirmative to the extent of any professional group differing from a random selection of the population. Additionally, we have drawn attention to our findings on the BEM measure which might suggest that teachers might be somewhat more androgynous in social roles than other groups, perhaps reflecting their focus of interest in teaching. Such differences are only



of import, however, to the extent of explaining any influence on our results. Finlay-Jones (1986) studied more than 2000 school teachers (of all ages) in Western Australia and established that 17% rated as having severe psychological distress as compared to a 9% rate for the general population, but that neither age nor sex had any influences on rates. Thus, teachers may have higher levels of psychological morbidity (for reasons that precede or are a consequence of a teaching career) and male and female teachers may well be expected to record similar rates of psychiatric morbidity over time. Whether the latter is a reflection of homogeneity of work roles, as suggested by Jenkins (1985) for British civil servants, or is more intrinsic to teachers, cannot be answered. The relevant question, however, is: even if teachers are so idiosyncratic as to have similar rates of psychological morbidity, how can any biological factor imputed as determining sex differences in the general population be over-ridden in such a select group? The most parsimonious explanation for the absence of a sex difference is that social role divergence for our cohort has been slight (against expectation), suggesting the key relevance of social factors as determining any sex difference found in general population studies. The importance of a negative finding, if confirmed in other studies, is that primary biological determinants could be dismissed, to allow a more refined set of possible determinants to be examined. In undertaking future studies there are clearly design problems. A null hypothesis cannot be proved, although it may be supported. While a similar study of a general population group would appear appropriate, there is a clear paradox in that similar attempts to control psychosocial variables in such a sample would ensure that the group is no longer representative of the general population. Statistical control of putative determinants in such a study may provide some useful information but we would argue that our present cohort design would be even more useful in a sample initially matched, and which subsequently diverged substantially in terms of social role variables, to test the critical hypothesis that change in depression rate covaries with change in social role. Such a situation may, of course, occur in our current cohort as we review members over time.

This research was assisted by funds provided by the New South Wales Institute of Psychiatry and the NH & MRC. We thank Sydney Teachers' College (University of Sydney) and the teachers for assistance in conducting the study, Dusan Hadzi-Pavlovic, Linda Hayward and Robin Curtain for data analyses, Rita Brown and Sandra Evans for secretarial assistance, and the following for comments on the paper-in-preparation: Gavin Andrews, Bryanne Barnett, Kerrie Eysers, Scott Henderson and Chris Tennant.

## REFERENCES

- American Psychiatric Association, (1980). *Diagnostic and Statistical Manual of Mental Disorders*, third edn. APA: Washington, DC.
- Angst, J. & Dobler-Mikola, A. (1984). The definition of depression. *Journal of Psychiatric Research* 18, 401-406.
- Anthony, J. C., Folstein, M., Romanoski, A. J., Von Korff, M. R., Nestadt, G. R., Chahal, R., Merchant, A., Brown, C. H., Shapiro, S., Kramer, M. & Gruenberg, E. M. (1985). Comparison of the lay Diagnostic Interview Schedule and a standardized psychiatric diagnosis: experience in Eastern Baltimore. *Archives of General Psychiatry* 42, 667-675.
- Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology* 42, 155-162.
- Blatt, S. J., D'Affili, J. P. & Quinlan, D. M. (1975). Depressive Experiences Questionnaire. Unpublished report, Department of Psychology, Yale University.
- Boyd, J. H. & Weissman, M. M. (1981). Epidemiology of affective disorders: a reexamination and future directions. *Archives of General Psychiatry* 38, 1039-1046.
- Briscoe, M. (1982). Sex differences in psychological well-being. *Psychological Medicine*. Monograph Supplement 1.
- Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*, third edn. Academic Press: New York.
- Congalton, A. A. (1969). *Status and Prestige in Australia*. Cheshire Publishing: Melbourne.
- Costello, C. G. & Comrey, A. L. (1967). Scales for measuring depression and anxiety. *Journal of Psychology* 66, 303-313.
- Endicott, J. & Spitzer, R. L. (1978). A diagnostic interview: the Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry* 35, 773-782.
- Eysenck, H. J. & Eysenck, S. B. (1964). *Manual of the Eysenck Personality Inventory*. Hodder and Stoughton: London.
- Finlay-Jones, R. (1986). Factors in the teaching environment associated with severe psychological distress among school teachers. *Australian and New Zealand Journal of Psychiatry* 20, 304-313.
- Folstein, M. F., Romanoski, A., Nestadt, G., Chahal, R., Merchant, A., Shapiro, S., Kramer, M., Anthony, J., Gruenberg, E. M. & McHugh, P. R. (1985). Brief report on the clinical appraisal of the Diagnostic Interview Schedule carried out at the Johns Hopkins site of the Epidemiological Catchment Area programme of the NIMH. *Psychological Medicine* 15, 809-814.
- Goldberg, D. P., Cooper, B. & Eastwood, M. R. (1970). A standardised psychiatric interview suitable for use in community surveys. *British Journal of Preventive and Social Medicine* 24, 18-23.
- Hagnell, O., Lanke, J., Rorsman, B. & Ojesjo, L. (1982). Are we entering an age of melancholy? Depressive illnesses in a prospective epidemiologic study over 25 years - the Lundby study. *Sweden Psychological Medicine* 12, 279-289.
- Hammen, C. L. & Padesky, C. A. (1977). Sex differences in the expression of depressive responses on the Beck Depression Inventory. *Journal of Abnormal Psychology* 86, 609-614.
- Helzer, J. E., Robins, L. N., McEvoy, L. T., Spitznagel, E. L., Stoltzman, R. K., Farmer, A. & Brockington, I. F. (1985). A comparison of clinical and Diagnostic Interview Schedule data.

- noses: Physician reexamination of lay-interviewed cases in the general population. *Archives of General Psychiatry* 42, 657-666.
- Hirschfeld, R. M. A., Klerman, G. L., Gough, H. G., Barrett, J., Kordins, J. & Chodoff, P. (1977). A measure of interpersonal dependency. *Journal of Personal Assessment* 4, 610-618.
- Jenkins, R. (1985). Sex differences in minor psychiatric morbidity. *Psychological Medicine Monograph Supplement* 7.
- Jorm, A. F. (1987). Sex and age differences in depression: A quantitative synthesis of published research. *Australian and New Zealand Journal of Psychiatry* 21, 46-53.
- Katz, R. & McGuffin, P. (1987) Neuroticism in family depression. *Psychological Medicine* 17, 155-161.
- Leckman, J. R., Sholomskas, D., Thompson, D., Belanger, A. & Weissman, M. M. (1982). Best estimate of lifetime prevalence diagnosis: a methodological study. *Archives of General Psychiatry* 39, 879-883.
- Murphy, J. M. (1986). Trends in depression and anxiety: Men and women. *Acta Psychiatrica Scandinavica* 73, 113-127.
- Oliver, J. M. & Simmons, M. E. (1985). Affective disorders and depression as measured by the Diagnostic Interview Schedule and the Beck Depression Inventory in an unselected adult population. *Journal of Clinical Psychology* 41, 469-476.
- Parker, G. (1979). Sex differences in non-clinical depression. *Australian and New Zealand Journal of Psychiatry* 13, 127-132.
- Parker, G. (1987). Are the lifetime prevalence estimates in the ECA study accurate? *Psychological Medicine* 17, 275-282.
- Regier, D. A., Myers, J. K., Dramer, M., Robins, L. H., Blazer, D. G., Hough, R. L., Eaton, W. W. & Locke, B. Z. (1984). The NIMH Epidemiologic Catchment Area program: historical context, major objectives and study population characteristics. *Archives of General Psychiatry* 41, 934-941.
- Reich, T., Rice, J., Andreasen, N. & Clayton, P. (1980). A preliminary analysis of the segregation distribution of primary major depressive disorder. *Psychiatric Bulletin* 16, 34-36.
- Renwick, P. A. & Lawler, E. E. (1978). Special report on work papers. *Psychology Today* 11, 53-65.
- Robins, L. N., Helzer, J. E., Croughan, J. & Ratcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics and validity. *Archives of General Psychiatry* 38, 381-389.
- Robins, L. N., Helzer, J. E., Weissman, M. M., Orvaschel, H., Gruenberg, G., Burke, J. D. & Regier, D. A. (1984). Lifetime prevalence of specific psychiatric disorders in three sites. *Archives of General Psychiatry* 41, 949-958.
- Romans-Clarkson, S. E., Walton, V. A., Herbison, G. P. & Mullen, P. (1988). A study of women who refused to participate in a community survey of psychiatric disorder. *Australian and New Zealand Journal of Psychiatry* 22, 19-29.
- Rosenberg, M. (1965). *The Measurement of Self-Esteem, Society and the Adolescent Self-Image*. Princeton University Press: Princeton, New Jersey.
- Spitzer, R. L., Endicott, J. & Robins, E. (1981). Research Diagnostic Criteria (RDC) for a Selected Group of Functional Disorders. Updated Third Edition (6/80). Typescript obtained from New York State Psychiatric Institute (Dr Endicott).
- Weissman, M. M. & Klerman, G. L. (1977). Sex differences and the epidemiology of depression. *Archives of General Psychiatry* 34, 98-111.
- Weissman, M. M. & Myers, J. K. (1978). Affective disorders in a US urban community. The use of Research Diagnostic Criteria in an epidemiological survey. *Archives of General Psychiatry* 35, 1304-1311.
- Weissman, M. M., Leaf, P. J., Holzer, C. E., Myers, J. K. & Tischler, G. L. (1984). The epidemiology of depression: an update on sex differences in rates. *Journal of Affective Disorders* 7, 179-188.
- Wilson, P. H. (1979). Behavioural and Pharmacological Treatment of Depression. Unpublished PhD thesis. University of New South Wales.

## The development of a measure of intimate bonds

KAY WILHELM<sup>1</sup> AND GORDON PARKER

*From the School of Psychiatry, University of New South Wales, Australia*

**SYNOPSIS** This paper discusses the relevance of assessing the nature of intimate relationships and reports on the development of such an instrument. The Intimate Bond Measure (IBM) is a self-report measure assessing two key underlying dimensions, care and control. Its properties are assessed in separate studies, establishing its high test-retest reliability, the homogeneous nature of the isolated dimensions, its insensitivity to broad socio-demographic influences and its minimal sensitivity to depressed mood state. Support for its validity, in terms of both perceived and actual characteristics of care and control, is demonstrated. It provides a simple and efficient measure of central constructs underlying intimate relationships, and is of potential use in studies attempting to assess the relevance of intimate relationships to the onset and course of psychiatric disorders.

### INTRODUCTION

In this paper we consider several measures of 'intimacy' and 'adjustment' currently being used to describe marital or other intimate relationships. We then define key constructs underlying such relationships, and describe the development of a self-report measure quantifying the derived constructs.

Our inquiry reflects current research interest in the relevance of social support systems and, in particular, intimate relationships in influencing an individual's resilience to adversity and the onset of neurotic decompensation. Henderson *et al.* (1981) established that the perceived adequacy was more important than quantitative aspects of the social support network in neurotic subjects, and their research highlights the need to examine subjective requirements of the individual when assessing social networks.

The importance of lack of intimacy as a risk factor affecting the onset and course of neurotic disorder has been suggested in a number of studies. Brown & Harris (1978) isolated four vulnerability factors to depression in a non-clinical female sample, one being the lack of a confiding intimate relationship. In a replication study in Alberta, Costello (1982) examined risk factors for depression in a female sample, and reported that a lack of intimacy with spouse, cohabitant or boyfriend increased the risk of

depression. In fact, Harris & Brown (1985) note that an intimate confiding relationship with a spouse has been described as protective against depression in nine of the ten relevant cross-sectional studies of female subjects.

The reparative capacity of social support has been documented. Quinton *et al.* (1984) noted that, while the selection of one's mate is influenced by one's own experience, selection of a non-deviant spouse who provides emotional support and good living conditions produces a significant protective effect against further emotional difficulties, even in a setting of adverse early experience. Parker & Hadzi-Pavlovic (1984) drew a similar conclusion after studying women bereaved of mothers early in life. In that study, any diathesis to depression established by earlier parenting appeared capable of significant modification by characteristics of the spouses.

Various strategies have been used to measure the quality of intimate relationships. The Interview Schedule for Social Interaction (ISSI), a semi-structured interview developed by Henderson *et al.* (1981), is a global measure of social support, but has been judged as limited in its capacity to measure confiding relationships, as only one question addresses that issue (O'Connor & Brown, 1984). Brown & Harris (1978) developed a semi-structured interview approach to assess psychosocial factors, to enquire specifically into the presence of a confidant to whom the subject might turn to discuss a problem, and the frequency of such contact. This approach

<sup>1</sup> Address for correspondence: Dr K. Wilhelm, School of Psychiatry, Prince Henry Hospital, Little Bay, 2036, Australia.

was modified by Costello (1982), who separated the question concerning intimacy with spouse from that assessing intimacy in confidants. The validity and utility of these approaches remain to be further established.

Concurrent with this increasing sophistication in investigation of social networks, there have been attempts to classify and measure dimensions of marital relationships, including marital 'happiness', 'success', 'stability' and 'adjustment'. The Locke-Wallace scale (1959) grouped key items from existing scales in an attempt to measure and predict marital adjustment. The scale was influential for the next two decades but has been criticized as being partly a measure of conventionality and social desirability. Spanier (1976) developed a widely used 'Dyadic Adjustment Scale' which assesses satisfaction, cohesion and affectional issues. It focuses on 'adjustment' as a general indicator of marital quality and the test is offered as an aid to uncovering problem areas in close relationships.

A further impetus to the development of measures of marital function was provided by British studies examining the relationship between poor marital quality and neuroticism (Pond *et al.* 1963), and the interaction of spouses with each other where the husband was a designated patient with a neurotic disorder (Kreitman, 1964). Kreitman *et al.* (1971) rated marital interaction in terms of 'assertiveness' and 'affection' and considered these to be independent constructs. Ryle (1966) developed a Marital Patterns Test (MPT) to quantify spouses' experience of affection given and received, and of domination/submission. The measure was developed initially to address a specific research question concerning the relationship of marital support and marriages involving a neurotic spouse. There have been problems with the selection of items in each dimension; Birtchnell (1985) has commented on the contamination of 'affection' items by 'domination' items, and overall there have been problems with the validity of the 'domination' items. However, the MPT has been subjected to ongoing validity studies (Heins & Yelland, 1981) and refinement (Scott-Heyes, 1982), and it has continued to generate interest, most recently by Birtchnell (1985). Shutz (1966) had earlier hypothesized three dimensions of interpersonal behaviour, namely 'inclusion', 'control' and

'affection', and had also measured these dimensions in terms of 'expressed' and 'wanted' behaviour, giving six subscales on his measure - the Fundamental Interpersonal Relations Orientation inventory (FIRO-B). The measure was designed to cover all interpersonal situations, but has had some use in marital assessment. Later, the Marital Attitudes Evaluations (MATE) was devised (Schutz, 1967). The measure has not been widely used and the inclusion of such items as 'I want my spouse to spend more time with me and give me more attention' and 'I want my spouse to allow me more freedom and think more for myself' is likely to be confounded by issues of social desirability and plaintive set.

The Maudsley Marital Questionnaire (MMQ) is a 20-item measure (Crowe, 1978) derived from the Structured and Scaled Interview to Assess Maladjustment (SSIAM) (Gurland *et al.* 1972). The MMQ was first used to assess progress in couples involved in conjoint marital therapy. Factor analysis of the MMQ generated three scales - 'marital adjustment', 'sexual adjustment' and 'general life adjustment'. Arrindell has examined the MMQ critically, undertaken further reliability and validity studies, and altered some items. The original 40-item version has now become a 20-item self-report scale (Arrindell *et al.* 1983*a,b*; Arrindell & Schaap, 1985). The measure has been used to investigate the effect of marital adjustment on treatment outcome for agoraphobia (Monteiro *et al.* 1985), and for phobic and obsessional disorders (Cobb *et al.* 1980). The MMQ is rightly less ambitious in its overall scope than the Wallace Locke Measure, but quantifies satisfaction and adjustment in the three defined areas. Arrindell has suggested that the MMQ would require a measure of intimacy to ensure a complete assessment of marital quality (Arrindell & Schaap, 1985).

Two questionnaires have been developed to measure the construct of intimacy. Waring's group has devised a structured interview, the Victoria Hospital Intimacy Interview (VHII), and a self-report scale, the Waring Intimacy Questionnaire (WIQ), with 'intimacy' being defined on eight subscales in each version (Waring *et al.* 1981; Waring & Reddon, 1983). Schaefer & Olson (1981) developed the Personal Assessment of Intimacy in Relationships

inventory (PAIR), an instrument which examines five components of perceived and expected intimacy exchange between partners. The authors make the point that intimacy is a process occurring over time, and that individuals have differing needs for intimacy, which they hold 'is never complete or fully accomplished'.

In our study we sought instead to define the key constructs underlying intimate relationships, commencing with a heterogeneous collection of items rather than pre-conceived constructs. The measures developed to date have generated items to reflect preconceived constructs in intimate relationships, which their authors felt were likely to be predictive of neuroticism (e.g. Ryle's MPT), or outcome after a therapeutic intervention (e.g. Crowe's MMQ), or to identify problem areas and marital satisfaction (e.g. Locke-Wallace, Spanier's DAS).

Our task was both more fundamental and ambitious than previous studies, in that we sought to identify the elements that define an intimate relationship. The initial impetus to development of the measure came from Hinde's plea for a sound basis of description and classification of interpersonal relationships (Hinde, 1979). He stated: 'It would in theory be possible to study a wide range of relationships, measure many of their aspects and then reduce the data by factor analysis or some comparable technique to a limited number of dimensions. A number of such attempts have been made, and in many cases the data have been successfully reduced to three principal dimensions, which approximate in everyday terms to dominance/subordinance, love/hate and involvement/detachment. But however many measures are used initially, some selection is necessarily involved, and the factors extracted from the analysis are inevitably influenced by that initial selection of data.' This approach also has an advantage in research terms, as too often associations between predictor and outcome variables have been spurious because the variables have not been independent of each other. Therefore, the use of a measure of marital satisfaction (as a predictor variable) to measure depression (as an outcome variable) leads to contamination of both sets of variables by a component of dissatisfaction. An example was mentioned previously in relation to the MATE scale (Schutz, 1967). Finally, we sought to develop a

self-report scale of equal relevance to both sexes which was easy to administer and score.

## METHODS

### Test construction and initial sample

Items were generated from a literature review and from interviews of married subjects who were asked to describe the behaviours and perceived attitudes of their spouse. The item pool was reduced to 83 by removal of synonymous items, those found to be ambiguous in pilot studies with volunteers, and items that appeared biased towards either sex.

As we sought to define items of general relevance, we elected not to select a sample of psychiatric patients. Principally for ease of data collection, we decided to obtain data mainly from general practice attenders. While such groups have been shown in many studies to have levels of psychological morbidity higher than in the general population and thus not indicative of the population at large, our previous experience suggested such a group would be appropriate to our task. The group of subjects was recruited from practices spread widely over the metropolitan area to allow a range of social class and life style. Respondents were routine attenders of 11 general practices and one optometry service in the Sydney area, and the subjects were requested by practice secretaries to complete the questionnaire anonymously and to then 'post' it in collection box placed in the waiting room. Secretaries screened subjects according to the following inclusion criteria: married subjects, between 20 and 65 years of age, having a reasonable knowledge of English, with no evidence of dementia or psychosis, and not obviously affected by drug or alcohol intoxication or by severe illness. These exclusion criteria were used as well for all the subsequent studies reported in this paper. Each subject was asked to score their partner's attitudes and behaviours in recent times on a four-point Likert-type scale (see Appendix 1). Additionally, subjects were asked to record their age, sex and occupation, the last being to assess socioeconomic status on the seven-point Congalton (1969) scale.

Forty-four of the 288 forms returned were discarded because of incomplete responses. The remaining forms (returned by 148 females and

96 males) generated the data for analysis. The mean age of the sample members was 39.7 (S.D. 12.8) years.

Principal components analysis was used with the Statistical Package for the Social Services (SPSS) program (Nie *et al.* 1975), and an oblique rotation selected. The first analysis incorporated all 83 items, and suggested 32 items to be weak discriminators with a poor distribution of responses. A second analysis entered the remaining 51 items. The unlimited factor solution was inspected, as well as imposed two-factor to six-factor pattern matrix solutions. The two-factor solution appeared the most interpretable and we therefore used this solution to generate the two subscales; these we labelled 'care' and 'control' and, after examining factor scores and communalities, we limited them to 12 items each. Factors 1 and 2 were weakly associated ( $-0.36$ ), as were the derived scores (see Appendix 1) on the total 'care' and 'control' scales ( $-0.45$ ), suggesting that high 'care' is intrinsically associated with less 'control'. Table 1 reports the factor loadings for the 24 items contributing to the final scales for the total sample, as well as the factor loadings derived separately for the males and for the females. The rank order of factor loadings suggested only a few sex effects. For instance, in relation to the care scale, females were somewhat more likely to rate their male partners as 'considerate' and 'understanding' (items 1 and 10), while males were somewhat more likely to rate their female partners as

'physically gentle and considerate' (item 13) and to make them 'feel needed' (item 14). By contrast, the rank order of variables on the control scale was rather similar for each sex.

Table 2 reports data from our principal sample. As the minimum score for each subscale is 0, and the maximum 36, the mean 'care' scale scores suggest skewing while scores for the 'control' scale are more normally distributed.

The influences of subjects' age, sex and social class on scale scores were examined by univariate analyses and subsequently by regression analyses, but no significant associations were demonstrated. The three socio-demographic variables accounted for only 1.9% of the variance in care scores and 3.0% of the variance in control scores. Thus the scales appear quite insensitive to broad socio-demographic influences.

We next assessed the following properties of the derived measure in a series of studies: internal consistency, test-retest reliability, effect of depressed mood on scale scores, and validity as a measure of both 'perceived' and 'actual' partner characteristics.

### Reliability studies

Two aspects of reliability were assessed: internal consistency and test-retest reliability. Internal consistency was extremely high in our principal sample, with Cronbach's alpha being 0.94 for the care scale and 0.89 for the control scale (Cronbach, 1951). Such results suggest that we have derived homogeneous dimensions. To assess test-retest reliability, the measure was completed by 28 normal volunteers (mean age 34 years, S.D. 6.3 years), with the interval between

Table 1. *Factor loadings for items contributing to the final items\* in the measure*

Care scale				Control scale			
Item no.	All	Male	Female	Item no.	All	Male	Female
16	0.81	0.80	0.82	11	0.75	0.84	0.71
5	0.79	0.82	0.79	12	0.71	0.62	0.77
4	0.78	0.72	0.80	7	0.69	0.66	0.71
1	0.75	0.67	0.78	17	0.68	0.77	0.71
18	0.74	0.71	0.76	20	0.67	0.66	0.74
21	0.73	0.72	0.72	15	0.63	0.63	0.62
10	0.72	0.66	0.72	2	0.61	0.52	0.65
8	0.72	0.71	0.71	3	0.60	0.45	0.67
14	0.71	0.83	0.64	19	0.59	0.45	0.68
23	0.70	0.71	0.69	9	0.59	0.54	0.62
24	0.70	0.70	0.69	6	0.57	0.55	0.59
13	0.70	0.79	0.65	22	0.55	0.61	0.51

\* Item numbers refer to those listed in Appendix 1.

Table 2. *Normative data from principal sample and validity studies groups*

	Care				Control		
	N	Mean	Median	S.D.	Mean	Median	S.D.
Principal sample							
Male	96	28.4	31.6	8.0	11.2	9.5	7.3
Female	148	27.1	29.2	8.3	9.6	7.1	8.3
Non-clinical sample (validity study 1)							
33	28.1	30.0	8.8	11.9	9.0	8.4	
Couples group (validity study 2)							
Male	25	23.6	24.0	8.6	13.8	11.0	8.6
Female	25	25.9	27.0	7.2	12.6	11.0	7.2

presentations varying from three to six weeks. Table 3 shows that the mean scores did not differ on the two occasions, while the reliability coefficients were very high (0.89 and 0.80,  $P < 0.001$  for both), supporting the reliability of the two derived scales.

As mood states are recognized as having the potential to influence self-report scores, the measure was completed by a group of depressives, both while depressed and after significant improvement. Severity of depression was monitored by either the Beck (Beck *et al.* 1961) or Zung (1965) measures, as subjects for this study were obtained from two other studies, variably using those two depression rating scales. While 35 subjects were entered into this study, only 30 were included in the final analyses as we imposed both a minimal improvement of ten units on either the Beck or Zung scale to ensure a distinct mood change and a clinical assessment of change from 'caseness' to 'non-caseness' in each subject. All 30 subjects had a clinical depressive disorder, most were in-patients at baseline assessment, and their diagnoses were broadly neurotic depression ( $N = 14$ ) and endogenous depression ( $N = 16$ ). Table 3 shows that there was a considerable and consistent decline in depression severity for the sample between occasions of testing, representing an improvement of 64% for the Beck scale and of 38% for the Zung scale, with parallel improvements in the level of depression. There was no significant alteration in the 'care' scale, but a small though significant change occurred in the 'control' scale.

### Validity studies

As the measure was designed to assess perceived characteristics, it appeared important to assess its validity, principally as a subjective measure. Thus a heterogeneous sample of 33 non-clinical volunteers and psychiatric patients was obtained (Validity study 1, Table 2), with subjects being interviewed by two raters who asked a series of predetermined questions aimed at eliciting the degree of 'care' and 'control' described by the subjects (e.g. 'How considerate is X towards you?' 'Who is likely to make decisions if there is a difference of opinion?'). These responses were rated on six-point ordinal scales. The first author (rater 'A') was a consistent rater of all 33 subjects, while the second rater ('B') was one of four psychologists who volunteered their time. The inter-rater reliability coefficients ('A' and 'B' scores intercorrelated) were assessed at 0.66 ( $P < 0.001$ ) for the 'care' dimension and 0.70 ( $P < 0.001$ ) for the 'control' dimension, suggesting moderate rater consensus in assessing these dimensions. After the interview, subjects completed the measure and scale scores and interview scores were then intercorrelated to assess the concurrent validity of the measure. 'Care' scale scores correlated 0.68 ( $P < 0.001$ ) with the level of 'care' judged by rater 'A' and 0.43 ( $P < 0.001$ ) with the 'care' judged by rater 'B'. 'Control' scale scores correlated 0.74 ( $P < 0.001$ ) with rater 'A' and 0.55 ( $P < 0.001$ ) with rater 'B' judgements of control at the interview. The higher coefficients returned against the first-author rater may reflect her constancy in assessing the content of the interview, acting to

Table 3. Normative data obtained in reliability studies

Study and scale	Baseline test (mean and s.d.)	Repeat test (mean and s.d.)	Correlation: time 1 with time 2	t-test time 1 – time 2 (two-tailed)
Test-retest study (non-clinical group)				
Care scale	28.4 (7.2)	28.0 (8.9)	0.89**	0.65
Control scale	8.2 (7.0)	7.8 (7.1)	0.80**	0.56
Mood state study (depressive sample)				
Care scale	24.2 (9.5)	25.6 (9.1)	0.92**	–1.5
Control scale	11.9 (8.7)	9.5 (7.8)	0.84**	2.6*
Depression levels				
Beck ( $N = 13$ )	22.2 ( 9.1)	8.1 (5.8)	0.67**	7.6**
Zung ( $N = 17$ )	55.6 (10.2)	34.6 (7.0)	0.80**	13.9**

\*  $P < 0.05$

\*\*  $P < 0.01$

reduce criterion variance, and her greater knowledge of the questionnaire and its putative dimensions.

While our preference was to develop a phenomenological measure of perceived characteristics, assessment of the degree to which the measure might reflect actual characteristics appeared important. To this end, we asked 25 couples who were engaged in marital therapy to complete the measure in relation to each other (Validity study 2, Table 2). One of two therapists (both social workers experienced in working with couples) made an objective assessment, using a six-point scale, of the interpersonal characteristics demonstrated by the husbands and wives towards each other during therapy. Their assessment used the same 'care' and 'control' probes as those used in the previous validity study.

If the scales are a valid measure of actual characteristics, scale scores should correlate with therapist ratings – assuming, of course, that the therapists were accurate raters and that the subjects behaved toward each other during therapy as they did generally. Raters' judgments of husbands' 'care' correlated 0.48 ( $P < 0.01$ ) with the wives' scores on our measure, while the equivalent examination for wives was 0.42 ( $P < 0.05$ ). Raters' judgements of husbands' 'control' correlated 0.51 ( $P < 0.005$ ) with the wives' scores on our measure, while the equivalent examination for wives returned a coefficient of 0.35 ( $P < 0.05$ ).

## DISCUSSION

We maintain that we have identified and defined two key dimensions underlying intimate interpersonal relationships. Our original set of additudinal and behavioural items was reduced, using principal components analysis, to suggest two source dimensions which we labelled 'care' and 'control'. The extent to which the measure might assess interpersonal relationships *in general* and not merely 'intimate' relationships cannot be answered, as our data and analyses were restricted to a respondent's 'partner'. In claiming to have developed a measure of 'intimate bonds' we are not claiming to measure intimacy itself (although the derived dimensions may contribute to the perceived presence or adequacy of intimacy). The word 'intimate' refers to our

intent to measure the contribution made by individuals' partners or their putative 'intimates'. The wording of some of the derived items (e.g. 'is physically gentle and considerate') would seem to preclude other less intense relationships. It would be of some interest to examine the extent to which individuals have any general tendency to 'care' or to 'control' in both intimate and other interpersonal relationships.

While the measure may well prove to be useful in measuring changes in intimate relationships and their consequences, it is important to know the extent to which scale scores remain constant for subjects whose general life situation is relatively stable, whether functional or dysfunctional. We therefore assessed the test-retest reliability of the measure over a reasonably short interval. In the non-clinical group described (see Table 3) the mean scores for each scale were stable over time and both correlation coefficients were high, strongly supporting the reliability of each scale, and indirectly supporting the validity of the measure.

A depressed mood has been frequently noted Paykel *et al.* 1969) to influence the perception or recall of experiences, and phenomenological or self-report measures are particularly susceptible to such effects. As it is likely that the measure will be used in case-control studies of depressed individuals to assess the relevance of intimate bonds, we assessed the degree to which scale scores might be modified by changes in mood state. Despite significant improvement in depression levels, the 'care' scale score was not significantly influenced by varying levels of depression. However, there was a change in perception of 'control' which reached significance and was in the direction that one would intuitively expect, namely, that the subjects, when depressed, tended to see their partner as being slightly more critical than when they were no longer depressed. This may be a reflection of change in the depressed person's behaviour or attribution by the depressed person of increased self-criticism to his/her partner, or it could reflect an actual change in spouse characteristics, as most subjects in this study were effectively rating their spouse before and after admission to hospital. However, the effect is not a large one. In relation to the stability of 'care' scores, the authors noted that subjects had an abiding



perception of whether or not they were cared for, even if they were depressed and not able to fully register the 'care' at that time. We conclude that any bias induced by a depressed mood is likely to be slight.

To assess the concurrent validity of the scales we undertook structured interviews with predetermined questions assessing perceived aspects of the partner's 'care' and 'control'. We established that the raters were in moderate agreement about both dimensions under investigation. Interrelation of interview scores returned by the one consistent rater with scale scores returned on our measure showed high agreement, supporting the concurrent validity of the derived measure. High correlations, however, might indicate that the subjects reported similarly in two contexts, and similar judgements might mean no more than a persistent response bias. It is clearly difficult to establish the validity of any phenomenological measure, so our analyses should be cautiously judged as suggestive of modest levels of validity.

Although our priority was to develop a self-report measure of constructs defining intimate relationships, it appeared important to make some estimate of the degree to which scale scores might reflect any 'objective reality'. Thus, we attempted to assess the validity of the measure in terms of its capacity to provide information about the 'actual' characteristics of the intimate relationship, assuming that there is likely to be some dissonance between 'actual' and 'perceived' characteristics. There are, as noted, considerable problems in attempting to validate any phenomenological construct and we are not aware of any technique that has resolved the intrinsic difficulties. The reader is referred to Spanier's (1979), Waring's (1985) and Birtchnell's (1985) reviews of the subject. Ryle (1966), in his validity study of the Marital Patterns Test, compared observations of a psychiatric social worker seeing the couple in question, his own observations of the couple in general practice and the couples' verbal self-rating with their scores on the Marital Patterns Test. Most validity studies have, however, concentrated on comparisons with other scales, not proceeding beyond concurrent validity.

A more commonly used strategy in extended validity studies is to compare scale scores against ratings provided by witnesses who are usually

family members or associates in a specific situation. There are distinct and rarely considered limitations to this approach. In an earlier study (Parker, 1983), using a cross-over sibling study design for the Parental Bonding Instrument, a clear limitation was noted, as respondents tended to score parents in a similar way, irrespective of whether they were asked to complete the questionnaire for themselves or on the basis of observations of their siblings. These findings suggest that there is no such reality as the 'objective' rater and that family members may be particularly likely to introduce their own subjective distortions, weakening any validity assessment using such 'independent' raters.

Crandall (1976) has reviewed studies attempting to validate self-report measures using ratings by others, and has noted rather low coefficients. We therefore attempted a modified strategy, by having couples engaged in marital therapy rate each other and be simultaneously rated by their therapist, who was blind to their scores. Clearly, there are limitations to this technique, as a number of biases may be contributed by the partners themselves (e.g. social desirability, defensiveness, need to hurt the other, and the degree to which key characteristics will actually be expressed in sessions), while the degree to which therapists may judge such characteristics objectively remains unclear. Nevertheless, the comparison of the therapists' judgement of each marital partner correlated moderately (mean = 0.44) with the marital partners' ratings of each other, a rather high level when compared with similar consensual estimates made by family witnesses (Crandall, 1976), and offers some support for the measure as one of the 'actual' characteristics of intimate relationships.

The care dimension, as defined by the twelve items, reflects care expressed emotionally as well as physically, with constructs of warmth, consideration, affection and companionship. These items are nearer to Birtchnell's concept of affection and less contaminated by those aspects of control which he noted in Ryle's 'affection' dimension (Birtchnell, 1985). The 'control' dimension suggests domination, intrusiveness, criticism, authoritarian attitudes and behaviours. The internal consistency analyses suggested that the two scales are highly homogeneous, an important property of the final measure. As the factor scores and the

scale scores on the two scales were negatively associated, it can be concluded that, in general terms, increasing 'control' in an intimate relationship is associated with less 'care'. We established that neither age, sex, nor social class had any influence on scale scores, another important property as it reduces the necessity to control rigorously for such variables in future case-control studies. The absence of any broad sex effect is somewhat surprising, but the absence of sex effects has also been noted by other writers (Waring & Reddon, 1983; Ryle, 1966) and it should be remembered that we had earlier deleted items which appeared at face value to be more specific or idiosyncratic to either sex.

Our earlier decision should be briefly considered. If we had left in overtly sex-specific items, then mean scale scores may have been influenced by sex of respondent and the measure might then be more sensitive to sexual differences in groups in any applied research. We preferred to develop a measure that would define central interpersonal dimensions, believing that such a measure would have greater utility in applied research. While we deleted a few obvious sex-specific items in the early work, we still anticipated a sex effect on derived scale scores. That it has not emerged is an intriguing finding.

While dimensions of 'care' and 'control' may be central to intimate relationships, intimacy itself is defined as 'a mutual needs' satisfaction' (Clinebell & Clinebell, 1970) and is in fact a bi-directional concept, with each partner measuring the intimate bond in terms of their own needs as well as the ingredients provided by the other. Our measure is principally uni-directional, measuring the degree to which the intimate is perceived as demonstrating certain attitudes or behaviours, although the form in which the items were phrased on the self-report scale must introduce a bi-directional component. Thus, while intimacy may theoretically be best effected by the combination of 'high care' and 'low control', the possibility should not be ignored

that some individuals may judge a lesser degree of care as satisfactory to their needs, while others may seek or require a moderate or high degree of 'control' from their intimate. Thus, while at this stage we would argue that we have defined key structural parameters in intimate interpersonal relationships, the judgement of 'intimacy' may require further assessment of the recipient's needs and satisfactions. Thus, our measure may well be complemented by one the dyadic adjustment scales which assess satisfaction-related issues.

Our derived Intimate Bond Measure (IBM) may be used to generate separate scale scores. Additionally, as for the Parental Bonding Instrument (Parker *et al.* 1979; Parker, 1983), the scales may be used together, allowing four broad styles of intimate relationships to be defined. Provisionally, we would label the 'high care-low control' quadrant as reflecting 'optimal intimacy', the 'high care-high control' quadrant as 'affectionate constraint', the 'low care-high control' quadrant as 'affectionless control' and the 'low care-low control' quadrant as defining an absence of intimacy; but, as noted above, the bi-directional nature of intimate bonds must be conceded and our labels run the risk of insufficiently acknowledging this issue. We suggest that the IBM may be a simple and efficient measure of constructs central to intimate relationships and be of use in risk and outcome studies.

Dr Wayne Hall assisted with the factor analysis. Dr Linda Hayward assisted the statistical analyses. Drs Lee and Robert Watson, Drs Eric and Anne Fisher, Dr Esther Kok, Dr Tony Wachter, Dr Barry Landa, Dr George Fromberg, Dr Bruce Andrews, Dr Reid Cameron, Dr. Andre Haski, Dr Gerard Barold, Dr Michael Armstrong, Dr David Fox and Associate Professor Brian Holden allowed access to their patients for collection of data. Wilma Sturgeon provided secretarial assistance. The New South Wales Institute of Psychiatry supported the first author as a Research Fellow. All are thanked with gratitude.

## APPENDIX 1. Instructions, items and scoring

This questionnaire lists some attitudes and behaviours which people reveal in their close relationships. Please judge your partner's attitudes and behaviour towards you in recent times and tick the most appropriate bracket for each item.

	Very true	Moderately true	Somewhat true	Not at all
1. Is very considerate of me	(3)*	(2)	(1)	(0)
2. Wants me to take his/her side in an argument	(III)†	(II)	(I)	(0)
3. Wants to know exactly what I'm doing and where I am	(III)	(II)	(I)	(0)
4. Is a good companion	(3)	(2)	(1)	(0)
5. Is affectionate to me	(3)	(2)	(1)	(0)
6. Is clearly hurt if I don't accept his/her views	(III)	(II)	(I)	(0)
7. Tends to try and change me	(III)	(II)	(I)	(0)
8. Confides closely in me	(3)	(2)	(1)	(0)
9. Tends to criticize me over small issues	(III)	(II)	(I)	(0)
10. Understands my problems and worries	(3)	(2)	(1)	(0)
11. Tends to order me about	(III)	(II)	(I)	(0)
12. Insists I do exactly as I'm told	(III)	(II)	(I)	(0)
13. Is physically gentle and considerate	(3)	(2)	(1)	(0)
14. Makes me feel needed	(3)	(2)	(1)	(0)
15. Wants me to change in small ways	(III)	(II)	(I)	(0)
16. Is very loving to me	(3)	(2)	(1)	(0)
17. Seeks to dominate me	(III)	(II)	(I)	(0)
18. Is fun to be with	(3)	(2)	(1)	(0)
19. Wants to change me in big ways	(III)	(II)	(I)	(0)
20. Tends to control everything I do	(III)	(II)	(I)	(0)
21. Show his/her appreciation of me	(3)	(2)	(1)	(0)
22. Is critical of me in private	(III)	(II)	(I)	(0)
23. Is gentle and kind to me	(3)	(2)	(1)	(0)
24. Speaks to me in a warm and friendly voice	(3)	(2)	(1)	(0)

\* 'Care' scale, Arabic numerals. † 'Control' scale, Roman numerals.

## REFERENCES

- Arindell, W. A. & Schaap, C. (1985). The Maudsley Marital Questionnaire (MMQ): an extension of its construct validity. *British Journal of Psychiatry* 147, 295-299.
- Arindell, W. A., Bodens, W. & Lambert, H. (1983a). On the psychometric properties of the Maudsley Marital Questionnaire (MMQ): evaluation of self-ratings in distressed and 'normal' volunteer couples based on the Dutch version. *Personality and Individual Differences* 4, 293-306.
- Arindell, W. A., Emmelkamp, P. M. G. & Bast, S. (1983b). The Maudsley Marital Questionnaire (MMQ): a further step towards its validation. *Personality and Individual Differences* 4, 457-464.
- Beck, A. T., Ward, C. M., Mendelson, M., Mock, J. E., Erbaum, T. & Waugh, J. K. (1961). An inventory for measuring depression. *Archives of General Psychiatry* 4, 561-671.
- Burchnell, J. (1985). The relationship between scores on Ryle's Marital Patterns Test and independent ratings of marital quality. *British Journal of Psychiatry* 146, 638-644.
- Brown, G. & Harris, T. (1978). *Social Origins of Depression. A Study of Psychiatric Disorder in Women*. Tavistock Publications: London.
- Clinebell, H. J. & Clinebell, C. M. (1970). *The Intimate Marriage*. Harper and Row: New York.
- Cobb, J. P., McDonald, R., Marks, I. & Stern, R. (1980). Marital versus exposure therapy - psychological treatment of co-existing marital and phobic-obsessive problems. *Behavior Analysis and Modification* 4, 3-16.
- Congalton, A. A. (1969). Status ranking list of occupations in Australia, Appendix B to *Status and Prestige in Australia*. F. W. Cheshire: Melbourne.
- Costello, C. G. (1982). Social factors associated with depression: a retrospective community study. *Psychological Medicine* 12, 329-339.
- Crandall, R. (1976). Validation of self-report measures using ratings by others. *Social Methods and Research* 4, 380-400.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297-334.
- Crowe, M. T. (1978). Conjoint marital therapy: a controlled outcome study. *Psychological Medicine* 8, 623-636.
- Gurland, B. J., Yorkston, N. J., Stone, A. R., Frank, J. D. & Fliess, L. (1972). The structured and scaled interview to assess maladjustment (SSIAM). I. Description, rationale and development. *Archives of General Psychiatry* 27, 259-263.
- Harris, T. & Brown, G. W. (1985). Interpreting data in aetiological studies of affective disorder: some pitfalls and ambiguities. *British Journal of Psychiatry* 147, 5-15.
- Heins, T. J. & Yellow, J. H. (1981). Validity studies on the Ryle Marital Patterns Test. *British Journal of Medical Psychology* 54, 51-58.
- Henderson, A. S., Byrne, D. G., & Duncan-Jones, P. (1981). *Neurosis and The Social Environment*. Academic Press: Sydney.
- Hinde, R. A. (1979). *Towards Understanding Relationships*. Academic Press: London.
- Kreitman, N. (1964). The patient's spouse. *British Journal of Psychiatry* 110, 159-173.

- Kreitman, N., Collins, J., Nelson, B. & Troop, J. (1971). Neurosis and marital interaction. IV. Manifest psychological interaction. *British Journal of Psychiatry* **117**, 33-46.
- Locke, H. J. & Wallace, K. L. (1959). Short marital-adjustment and prediction tests: their reliability and validity. *Marriage and Family Living* **21**, 251-255.
- Monteiro, W., Marks, M. & Ramm, E. (1985). Marital adjustment and treatment outcome in agoraphobia. *British Journal of Psychiatry* **146**, 383-390.
- Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. & Bent, D. M. (1975). *Statistical Package for the Social Sciences*. McGraw-Hill: New York.
- O'Connor, P. & Brown, G. W. (1984). Supportive relationships: fact or fancy? *Journal of Social and Personal Relationships* **1**, 159-175.
- Parker, G. (1983). *Parental Overprotection: A Risk Factor in Psychosocial Development*. Grune and Stratton: New York.
- Parker, G. & Hadzi-Pavlovic, D. (1984). Modification of levels of depression in mother-bereaved women by parental and marital relationships. *Psychological Medicine* **14**, 125-135.
- Parker, G., Tupling, M. & Brown, L. B. (1979). A parental bonding instrument. *British Journal of Medical Psychology* **52**, 1-10.
- Paykel, E. S., Myers, J. K., Dienelt, M. N., Klerman, G. L., Lindenthal, J. J. & Pepper, M. P. (1969). Life events and depression: a controlled study. *Archives of General Psychiatry* **21**, 753-760.
- Pond, D. A., Ryle, A. & Hamilton, M. (1963). Marriage and neurosis in a working class population. *British Journal of Psychiatry* **109**, 592-598.
- Quinton, D., Rutter, M. & Liddle, C. (1984). Institutional rearing, parenting difficulties and marital support. *Psychological Medicine* **14**, 107-124.
- Ryle, A. (1966). A marital patterns test for use in psychiatric research. *British Journal of Psychiatry* **123**, 535-541.
- Schaefer, M. T. & Olsen, D. H. (1981). Assessing intimacy: the PAIR inventory. *Journal of Marital and Family Therapy* **7**, 47-60.
- Schutz, William, C. (1966). *The Interpersonal Underworld. FIRO: A 3D Theory of Interpersonal Behaviour*. Science and Behaviour Books Inc: Palo Alto, California.
- Schutz, William, C. (1967). *The Firo Scales Manual*. Consulting Psychologists Press: Palo Alto, California.
- Scott-Heyes, G. (1982). Analysis and revision of Ryle's Marital Patterns Test. *British Journal of Medical Psychology* **55**, 67-75.
- Spanier, G. B. (1976). Measuring dyadic adjustment: new scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family* **38**, 15-28.
- Spanier, G. B. (1979). The measurement of marital quality. *Journal of Sex and Marital Therapy* **5**, 288-299.
- Waring, E. M. (1985). Measurement of intimacy: conceptual and methodological issues of studying close relationships. *Psychological Medicine* **15**, 9-14.
- Waring, E. M. & Reddon, J. R. (1983). The measurement of intimacy in marriage: The Waring Intimacy Questionnaire. *Journal of Clinical Psychology* **39**, 53-57.
- Waring, E. M., McElrath, D., Lefcoe, D. & Weisz, G. (1981). Dimensions of intimacy in marriage. *Psychiatry* **4** 471-474.
- Zung, W. W. K. (1965). A self-rating depression scale. *Archives of General Psychiatry* **12**, 63-70.

*Appendix IV*  
*Papers published on data*  
*from the thesis*

## **RELIABILITY OF THE PARENTAL BONDING INSTRUMENT AND INTIMATE BOND MEASURE SCALES**

Kay Wilhelm and Gordon Parker

---

**The long-term reliability of the Parental Bonding Instrument (PBI) and of the Intimate Bond Measure (IBM) are examined in a non-clinical group, with data being examined over eleven and five years for the two respective measures. Such reliability data are compared with reliability data on a number of personality measures within the same cohort. Results demonstrate considerable stability in the PBI over an extended period and moderate stability in IBM scores.**

**Australian and New Zealand Journal of Psychiatry 1990; 24:199-202**

The Parental Bonding Instrument (PBI) was developed to measure fundamental parental dimensions of care and protection (control) and to allow quantification of any parental contribution to subsequent psychiatric disorder. Reliability aspects (internal consistency and short-term test-retest reliability) were examined in the initial paper [1] and its validity, both as a measure of perceived and actual parenting, has been examined in several studies [2].

If the PBI is a reliable and valid measure of subjects' retrospective memories of their parents in their first sixteen years, then adult subjects should return consistent scores over time (ie show high test-retest reliability).

The short-term reliability of the PBI has been examined in several studies. In the initial development paper [1], test-retest reliability in a non-clinical sample was 0.76 for the care scale and 0.63 for the protection scale over a three-week interval. Subsequently, in a sample of depressives initially depressed and then

significantly improved, much higher correlation coefficients (ranging from 0.87 to 0.92) were returned over a nine-week interval [2]. The higher coefficients in this group, compared to the initial sample, were judged to reflect the greater motivation of patients (in comparison to volunteer or importuned non-clinical groups) to return questionnaire data conscientiously. Subsequently, in a US study [3] of depressed out-patients attending the Yale Depression Research unit, 48 depressives scored the PBI when depressed and some four-six weeks later when significantly improved. PBI scores showed no significant change over time and the coefficients of agreement ranged from 0.90 to 0.96 across the four scales, slightly superior to the Australian depressed sample. Test-retest reliability in a group of patients with schizophrenia has been examined [4] when the coefficients of agreement ranged from 0.58 to 0.77. This less impressive result was judged to be a reflection of the sample initially scoring the PBI shortly after admission to hospital with an exacerbation of their schizophrenia, with their judgement and ability to complete the self-report questionnaire impaired. This interpretation is supported by an American study [5] of 26 subjects with schizophrenia who completed the PBI form on two occasions a few weeks apart, with correlation coeffi-

cients ranging from 0.79 to 0.88, with this sample being distinguished by the sample being selected from those attending a community mental health centre, and not assessed during a relapse.

Medium-term reliability data have been provided [6], with correlation coefficients ranging from 0.79 to 0.81 on the PBI for a non-clinical U.S. sample tested seven months apart. In the only published study [7] of the long-term reliability of the PBI, women were studied in the post-partum period and then two to four years (mean 30 months) later. Only the maternal PBI form was completed by subjects, and the authors reported that PBI scores were "remarkably stable over time" for different sub-groups of mothers who were either depressed on both occasions of testing, initially depressed and then recovered, or not depressed on either occasion.

In this paper we examine the test-retest reliability of the PBI over a decade. The sample has been described elsewhere [8] and so details only will be summarized here. In September 1978, 380 students who had undertaken a basic Arts or Science university course, and who were then completing a one-year training programme at the Sydney Teachers' College were approached in class and invited to participate in a longitudinal study. While our key objective was to examine for sex differences in depressive experience over time, the students were not informed about the specific hypothesis, but were given details on the range of topics (including depression) and developmental issues that would be assessed longitudinally. Those subjects completed PBI data and 170 agreed to take part in the longitudinal study, and so formed the study cohort, then having a mean age of 23.1 years. Those taking part and those declining did not differ in PBI scores returned for each parent. Subsequently, we sought to interview the cohort serially, and self-report data were obtained from 164 in 1983 and 163 in 1988. On each occasion the subjects were requested to complete the orthodox PBI forms (assessing parenting over the first 16 years), allowing us to compare PBI data collected over extended periods.

The authors have also developed a measure of fundamental dimensions underlying adult intimate relationships [9]. The test-retest reliability of that self-report, the Intimate Bond Measure (or IBM) was assessed in the initial paper, with a non-clinical sample returning data on two occasions over a three-six week interval, with reliability coefficients being very high at 0.80 and 0.89. That measure was given to our

present cohort in 1983 and in 1988 and we now report the test-retest reliability over a five-year interval. It must be kept in mind, however, that subjects would not necessarily be scoring the same "intimate" on both occasions, so that we report consistency data for the whole sample and for a sub-sample of those who were married in 1988 and had rated the same "intimate" in 1983 - a fairer test of the measure's reliability. On each occasion, subjects were asked to score characteristics of the intimate "in recent times", the IBM being more a measure of state or current characteristics.

We also take the opportunity to report test-retest reliability over the same extended period for a number of other measures. We do that for several reasons. Firstly, such reports are rare and, more importantly, those data provide some basis for comparison against the PBI and IBM. It is generally suggested that personality is constant and we might therefore expect that high test-retest reliability would be demonstrated for personality measures and give a base quantitative estimate of reliability against which we could judge PBI data, in particular. That is, if personality is immutable, then self-report measures of personality should show a high level of constancy, being weakened only by response biases and state effects (eg depression) which are generally accepted to influence self-report scoring. Thus, we would expect that if the PBI is a reliable measure, reliability coefficients should be similar to those returned on personality measures. The personality measures considered were the Eysenck Personality Inventory neuroticism scale [10], the Rosenberg self-esteem scale, [11] the dependency scale from the Depressive Experiences Questionnaire [12], the Costello-Comrey trait depression scale [13] which was designed to measure a "person's tendency to experience a depressive mood", and the Bem sex role inventory [14] (with masculinity, femininity and social desirability sub-scales), the last being administered in 1983 and 1988 only. Finally, we report data on the Wilson-Lovibond state measure of depression [15] to again allow comparison against the "trait" measures, anticipating that much lower levels of agreement should be demonstrated on a state measure over time.

## Results

Table 1 reports the mean data returned for the several measures, and the level of constancy over time, with three intervals (1978-1983, 1983-1988, and

*Table 1. Consistency in scores examined over time*

	Mean score			Consistency coefficients (r)		
	1978 (A)	1983 (B)	1988 (C)	A with B	B with C	A with C
<b>PBI</b>						
Maternal care	26.3	26.2	26.3	0.72	0.82	0.63
Maternal protection	14.8	13.8	13.8	0.74	0.76	0.68
Paternal care	21.9	21.7	21.4	0.80	0.82	0.72
Paternal protection	13.0	12.1	11.9	0.69	0.67	0.56
<b>IBM</b>						
Intimate care (whole group)	N/A*	30.6	29.5	N/A	0.50	N/A
Intimate control (whole group)	N/A	6.7	6.9	N/A	0.49	N/A
Intimate care (sub-group)**	N/A	31.5	30.2	N/A	0.48	N/A
Intimate control (sub-group)	N/A	6.5	7.0	N/A	0.49	N/A
<b>Personality</b>						
Neuroticism	9.0	8.7	8.7	0.54	0.68	0.50
Self-esteem***	1.6	1.0	0.9	0.43	0.61	0.48
Dependency	52.6	52.8	53.7	0.64	0.64	0.55
Trait depression	31.3	31.2	29.4	0.64	0.65	0.46
<b>Sex role Inventory</b>						
Masculinity	N/A	4.6	4.7	N/A	0.56	N/A
Femininity	N/A	4.7	4.8	N/A	0.62	N/A
Social desirability	N/A	5.2	5.3	N/A	0.57	N/A
State depression	57.0	55.2	55.3	0.25	0.23	0.17

\* N/A = not assessed

\*\* Those nominating the same intimate in 1983 and 1988 (N=96)

\*\*\* Higher scores indicate lower self-esteem

1978-1988) being examined. Mean scale scores were generally stable for most measures apart from self-esteem, which showed a distinct improvement from 1978 to 1983 ( $t = 4.80$ ,  $P < 0.001$ ) and from 1978 to 1988 ( $t = 5.82$ ,  $P < 0.001$ ) but no change from 1983 to

1988 ( $t = 0.87$ , ns). In a similar, but less distinct fashion, depression scores decreased from 1978 to 1988, both on the trait ( $t = 2.03$ ,  $P < 0.05$ ) and state ( $t = 2.21$ ,  $P < 0.05$ ) measures.

Presumably because of the large sample size (for mean scores were very similar), two significant differences were established for the PBI and the IBM measures. Thus paternal protection scores dropped from 1978 to 1988 ( $t = 2.12$ ,  $p < 0.05$ ) while IBM care scores decreased from 1983 to 1988 ( $t = 2.86$ ,  $p < 0.05$ ). Scores were generally more consistent in the interval 1983-1988, when the cohort had left university and most were in full-time employment. The data for the 11-year interval (1978-1988) show slightly less consistency, presumably reflecting changes in attitudes over the lengthier assessment period. The test-retest consistency in PBI scores is extremely impressive for all three test intervals. Thus, the mean correlation coefficients were 0.74 (1978-1983), 0.77 (1983-1988) and 0.65 (1978-1988), contrasting with the four personality measures which returned mean correlations of 0.56, 0.64 and 0.50 respectively. Additionally, the correlation coefficients for the PBI were superior to each individual personality test. Stability in IBM scores (both for the whole sample and for the sub-group scoring the same partner) was moderate from 1983 to 1988 and clearly less stable than PBI scores, as might be anticipated for a state measure. Finally, the test-retest reliability of the state depression measure was low, as anticipated.

## Discussion

The sample allows a "best estimate" of reliability, in that it involved a non-clinical group, comprising volunteers who were prepared to take part in a longitudinal study and who were, at each review period, judged to be generous in giving their time for extended interviews and open in their discussion with the interviewers. Thus, we judge that sample members were likely to have completed self-report measures conscientiously and as accurately as possible. Additionally, being a non-clinical sample, it is unlikely that mood disturbance or related factors weakening reliability estimates would have been over-represented in the sample. The data set therefore offers a "best estimate" potential for any examination of reliability and it would be unlikely that non-volunteer or certain clinical groups would return such high levels of agreement. Thus, we acknowledge the unique characteristics of

the cohort but suggest that such a group is the appropriate one for such an examination.

The test-retest reliability of the PBI is clearly impressive, both intrinsically when the correlation coefficients are examined and, secondly, in comparison to the "personality" tests which we used as our comparative base. Jorm [16] considered the test-retest consistency of trait anxiety/neuroticism measures, and referenced work giving a correlation of 0.54 over the four-six years and 0.40 over 30 years, with lower levels for state measures of anxiety. Those data are compatible with our neuroticism score data, and therefore support the likely accuracy of the latter.

The test-retest data for the IBM are somewhat less impressive and, as we undertook a separate analysis on those in stable relationships with similar results, findings cannot reflect a sub-sample scoring different intimates over time. As the IBM is a measure of current intimate relationships, some change in the perception of the spouse or intimate would be anticipated over time, particularly in such a sample of young adults. While change in parenting over the years might also be theoretically anticipated, the PBI, by contrast, is designed deliberately to obtain an overall gestalt of the parent or "product moment of innumerable experiences" so that the instructions effectively force some overall judgement in an attempt to minimise variation at different developmental stages.

Each of the "personality" measures returned similar levels of consistency over time despite some of them (eg trait depression) conceivably being more likely to be influenced by mood state and therefore potentially unstable over an extended period. The extent to which they necessarily reflect intrinsic personality, however, cannot be addressed by such a study design.

We conclude then that the PBI is a highly reliable measure over an extended period, supporting its claim to be accurate measure of perceived parenting, and so useful in quantifying any parental risk to subsequent psychiatric disorder in adulthood.

## Acknowledgements

We thank our colleagues at the Mood Disorders Unit for comments on this paper, the NH & MRC and NSW Institute of Psychiatry for funding the cohort study, and Mrs Sandra Evans for typing of the manuscript.

## References

1. Parker G. A parental bonding instrument. *British Journal of Medical Psychology* 1979;52:1-10.
2. Parker G. Parental overprotection: a risk factor in psychosocial development. New York: Grune & Stratton, 1983.
3. Plantes MM, Prusoff BA, Brennan J, and Parker G. Parental representations of depressed outpatients from a U.S. sample. *Journal of Affective Disorders* 1988;15:149-155.
4. Parker G, Fairley M, Greenwood J, Jurd S, and Silove D. Parental representations of schizophrenics and their association with onset and course of schizophrenia. *British Journal of Psychiatry* 1982;141:573-581.
5. Warner R, and Atkinson M. The relationship between schizophrenic patients' perceptions of their parents and the course of their illness. *British Journal of Psychiatry* 1988;153:344-353.
6. Richman, JA, and Flaherty JA. Adult psychosocial assets and depressive mood over time: Effects of internalized childhood attachments. *Journal of Nervous and Mental Disease* 1987;175:703-712.
7. Gotlib IH, Mount JH, Cordy NI and Whiffen VE. Depression and perceptions of early parenting: A longitudinal investigation. *British Journal of Psychiatry* 1988;152:24-27.
8. Wilhelm K and Parker G. Are sex differences in depression really necessary? *Psychological Medicine* 1989;19:401-414.
9. Wilhelm K and Parker G. The development of a measure of intimate bonds. *Psychological Medicine* 1988;18:225-234.
10. Eysenck HJ and Eysenck SB. *Manual of the Eysenck Personality Inventory*. London: Hodder and Stoughton, 1964.
11. Rosenberg M. *The measurement of self-esteem. Society and the adolescent self-image*. New Jersey: Princeton University Press, 1965.
12. Blatt SJ, D'Affliti JP and Quinlan DM. *Depressive Experiences Questionnaire*. Unpublished report, Department of Psychology, Yale University, 1975.
13. Costello CG and Comrey AL. Scales for measuring depression and anxiety. *Journal of Psychology* 1967;66:303-313.
14. Bem SL. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology* 1974;42:155-162.
15. Wilson PH. *Behavioural and pharmacological treatment of depression*. Unpublished PhD thesis. University of New South Wales, 1979.
16. Jorm AF. Modifiability of trait anxiety and neuroticism: A meta-analysis of the literature. *Australian and New Zealand Journal of Psychiatry* 1989;23:21-29.