

## **INFORMATION TO USERS**

**This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.**

**The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.**

**In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.**

**Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.**

**Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.**

**ProQuest Information and Learning  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
800-521-0600**

**UMI<sup>®</sup>**



**Longitudinal Patterns of Depressive Symptoms in Midlife Women**

by

**Anne Marie Mariella**

**A dissertation submitted in partial fulfillment of the  
requirements for the degree of**

**Doctor of Philosophy**

**University of Washington**

**2001**

**Program Authorized to Offer Degree: Nursing**

**UMI Number: 3036501**

**Copyright 2001 by  
Mariella, Anne Marie**

**All rights reserved.**

**UMI<sup>®</sup>**

---

**UMI Microform 3036501**

**Copyright 2002 by ProQuest Information and Learning Company.  
All rights reserved. This microform edition is protected against  
unauthorized copying under Title 17, United States Code.**

---

**ProQuest Information and Learning Company  
300 North Zeeb Road  
P.O. Box 1346  
Ann Arbor, MI 48106-1346**

**© Copyright 2001**

**Anne Marie Mariella**

**Doctoral Dissertation**

**In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of the dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to Bell and Howell Information and Learning, 300 North Zeeb Road, P.O. Box 1346, Ann Arbor, MI 48106-1346, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."**

Signature   
Anne Marie Mariella

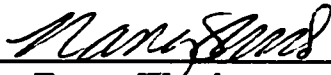
Date August 26, 2001

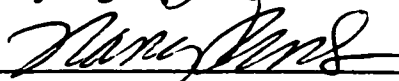
University of Washington  
Graduate School


This is to certify that I have examined this copy of a doctoral dissertation by

Anne Marie Mariella

and have found that it is complete and satisfactory in all respects,  
and that any and all revisions required by the final  
examining committee have been made.

Chair of Supervisory Committee:   
Nancy Fugate Woods

Reading Committee:   
Nancy Fugate Woods

  
Susan L. Woods

  
Pamela H. Mitchell

Date: August 27, 2001

**University of Washington**

**Abstract**

**Longitudinal Patterns Of  
Depressive Symptoms In Midlife Women  
by  
Anne Marie Mariella**

**Chair of Supervisory Committee  
Professor Nancy Fugate Woods  
Nursing**

**Many people assume that women in midlife, approaching and completing menopause, will experience increased depression. Yet most studies show the best predictor of depression in midlife is a history of previous depression. The aims of this study were to describe and characterize longitudinal patterns of depressive symptoms in midlife women before, during, and after the transition to menopause and to explore relationships between these patterns and menopausal transition stage (MTS), history of abuse/assault, and general life events and stressors.**

**Women were sampled from the Seattle Midlife Women's Health study (SMWHS), a descriptive longitudinal study of women who at entry to the study in 1990-92 were aged 35-55 years, resided in Seattle, spoke and read English, had at least some high school education, still had their uterus and at least one ovary, and had had a menstrual period within the previous 12 months. Depressive symptoms were measured approximately annually by the Center for Epidemiologic Studies-Depression scale (CES-D), with scores considered a continuous variable. For determination of longitudinal patterns, SMWHS participants had to have at least 6 (or 5 for hierarchical cluster**

analysis) and up to 10 annually repeated scores. Two cluster analyses and a phenomenological analysis were conducted on samples of 216, 249, and 205 women, respectively.

The important parameters of longitudinal patterns were level or severity of depressive symptoms (non-depressed, low, moderate, and high), duration of symptoms at that level (yearly or slower change), range and fluctuation (e.g., change in CES-D score >15 yearly), and trend over time (stable, increasing, or decreasing). The single largest group was stable non-depressed at all time points. MTS and life events were not related to CES-D score or longitudinal pattern. History of sexual abuse/assault was significantly related to a high chronic pattern of depressive symptoms (Pearson  $\chi^2 = 15.23$ ; df 6;  $p=.019$ ).

Clinical implications were recommendations to improve detection of history of depression and abuse/assault, and anticipatory guidance for long-term self-managed plans of care, regardless of MTS.

## TABLE OF CONTENTS

<b>List of Figures</b>		vi
<b>List of Tables</b>		ix
<b>Chapter 1:</b>	<b>Problem Statement</b>	1
	<b>Aims</b>	6
	<b>Study Questions</b>	6
<b>Chapter 2:</b>	<b>Background and Significance</b>	8
	<b>Definitions</b>	8
	<b>Severity</b>	9
	<b>Depressed Mood</b>	9
	<b>Depressive Symptoms</b>	9
	<b>Major Depressive Disorder</b>	10
	<b>Minor Depressive Disorder and</b>	
	<b>Subsyndromal Depression</b>	11
	<b>Dysthymic Disorder or Dysthymia</b>	11
	<b>Duration</b>	12
	<b>Course</b>	13
	<b>Causal Models of Depression</b>	15
	<b>Studies of the Course of Depression</b>	24
	<b>Longitudinal Studies of Depression and Menopause</b>	30
	<b>Samples</b>	32
	<b>Sampling Methods</b>	32
	<b>Sample Sizes</b>	35
	<b>Eligibility Criteria</b>	36
	<b>Sociocultural Characteristics of Samples</b>	38
	<b>Cohort Effects</b>	40
	<b>Measurement Tools</b>	42
	<b>Menopausal and Menstrual Status</b>	43
	<b>World Health Organization Definitions</b>	44
	<b>Evidence-based Definitions</b>	47
	<b>Comparisons of Definitions and Implications</b>	51
	<b>Preclimacteric</b>	53
	<b>Premenopausal</b>	53
	<b>Perimenopausal</b>	56
	<b>Postmenopausal</b>	59
	<b>Surgical Menopause</b>	60
	<b>Hormone Replacement Therapy</b>	61
	<b>Other Independent Variables</b>	63
	<b>Life Events and Stressors</b>	64

	Age .....	66
	Vasomotor Symptoms .....	67
	Marital Status .....	68
	Socioeconomic Status (SES) and Other Demographics.....	69
	Health Status .....	69
	Attitudes and Cultural Beliefs.....	70
	Hypothalamic-Pituitary-Ovarian Hormones .	70
	Designs .....	72
	Analysis Strategies and Techniques.....	74
	Norwegian Menopause Project .....	74
	Seattle Midlife Women’s Health Study .....	76
	University of Pittsburgh Healthy Women Study	76
	Massachusetts Women’s Health Study .....	77
	National Health Examination Follow-Up Study (NHEFS) .....	78
	Longitudinal Study of Women in Gothenburg	79
	South-East England Study.....	79
	Manitoba Project on Women and their Health in the Middle Years .....	80
	Melbourne Women’s Midlife Health Project	82
	Results .....	85
	Seattle Midlife Women’s Health Study .....	85
	University of Pittsburgh Healthy Women Study	86
	Massachusetts Women’s Health Study .....	87
	NHEFS .....	91
	Manitoba Project on Women and their Health in the Middle Years .....	92
	South-East England Study.....	94
	Melbourne Women’s Midlife Health Project	95
	Norwegian Menopause Project .....	96
	Longitudinal Study of Women in Gothenburg	96
	Summary .....	97
	Significance .....	101
Chapter 3:	Methods—Parent Study and General Methods.....	110
	Population .....	111
	Sampling Frame .....	111
	Sampling Procedure .....	112
	Contact Procedures.....	112
	Eligibility/Exclusion Criteria .....	112
	Sampling Results of Parent Study.....	113
	Tools and Measures .....	113
	Center for Epidemiologic Studies-Depression scale.	113

	Menopausal Transition Stage (MTS) .....	117
	Antidepressant and Tranquillizer Medications .....	119
	Study Protocol .....	121
	Sampling Frame of Current Study .....	122
	Measurement Procedures of Current Study .....	122
	Ethical Considerations.....	123
<b>Chapter 4:</b>	<b>Cluster Analyses .....</b>	<b>125</b>
	Aims .....	125
	Research Design .....	125
	Methods .....	126
	Review of Analysis Methods .....	127
	Sampling Procedure .....	130
	<b>Iterative or K-Cluster Analysis of Yearly CES-D Scores</b>	
	Plus SLOPE99.....	131
	Analytic Procedures .....	131
	Type of Cluster Analysis.....	131
	Number of Cluster Solutions.....	131
	Selection of Variables .....	131
	Results .....	133
	Sample.....	133
	Number of Cluster Solutions.....	133
	Clusters.....	134
	K-Cluster Summary.....	138
	K-Cluster Discussion.....	138
	<b>Hierarchical Agglomerative Cluster Analysis of .....</b>	
	Summary Measures .....	139
	Analytic Procedures .....	139
	Type of Cluster Analysis.....	139
	Number of Cluster Solutions.....	140
	Selection of Variables .....	140
	Method of Hierarchical Clustering.....	141
	Measure of Hierarchical Clustering .....	141
	Results .....	142
	Sample.....	142
	Number of Cluster Solutions.....	143
	Selection of Variables .....	143
	Method and Measure.....	144
	Clusters.....	144
	Hierarchical Cluster Summary .....	146
	Hierarchical Cluster Discussion .....	146
	Validity and Reliability .....	148
	<b>Predictors and Confounders .....</b>	<b>149</b>
	Menopausal Transition Stage (MTS) .....	150

	History of Assault/Abuse .....	151
	Antidepressant Drug Use .....	153
	Predictors and Confounders Summary.....	154
	Predictors and Confounders Discussion.....	155
<b>Chapter 5:</b>	<b>Phenomenological Analysis .....</b>	<b>188</b>
	<b>Aims .....</b>	<b>188</b>
	<b>Review of Analytic Method .....</b>	<b>188</b>
	<b>Research Design .....</b>	<b>192</b>
	<b>Methods .....</b>	<b>192</b>
	<b>Sampling Frame .....</b>	<b>192</b>
	<b>Sampling Procedure .....</b>	<b>192</b>
	<b>Tools and Measures.....</b>	<b>192</b>
	<b>Analytic Procedures .....</b>	<b>194</b>
	<b>Results .....</b>	<b>196</b>
	<b>Sample .....</b>	<b>196</b>
	<b>Longitudinal Patterns .....</b>	<b>196</b>
	<b>Waveform Patterns.....</b>	<b>196</b>
	<b>No Trend .....</b>	<b>200</b>
	<b>Rising .....</b>	<b>206</b>
	<b>Falling.....</b>	<b>207</b>
	<b>Grand W .....</b>	<b>209</b>
	<b>Non-waveform Patterns.....</b>	<b>210</b>
	<b>Stable.....</b>	<b>210</b>
	<b>Generally Linear Rising, Falling.....</b>	<b>212</b>
	<b>Blips .....</b>	<b>214</b>
	<b>Stairsteps .....</b>	<b>215</b>
	<b>Combination Patterns .....</b>	<b>217</b>
	<b>Scatterplots .....</b>	<b>219</b>
	<b>Aims .....</b>	<b>219</b>
	<b>Analytic Procedures .....</b>	<b>219</b>
	<b>Results .....</b>	<b>219</b>
	<b>Summary .....</b>	<b>221</b>
	<b>Discussion .....</b>	<b>223</b>
	<b>Interpretation of Findings.....</b>	<b>223</b>
	<b>Comparison with Other Studies .....</b>	<b>225</b>
	<b>Comparison of Phenomenological Results</b>	
	<b>to Cluster Results .....</b>	<b>228</b>
	<b>Limitations .....</b>	<b>229</b>
<b>Chapter 6:</b>	<b>General Discussion .....</b>	<b>254</b>
	<b>Interpretation of Findings.....</b>	<b>254</b>
	<b>Clinical Implications .....</b>	<b>258</b>

<b>Limitations</b> .....	<b>261</b>
<b>Implications for Further Research</b> .....	<b>265</b>
<b>Reference List:</b> .....	<b>267</b>
<b>Appendix A: Questionnaires and Data Collection Instruments</b> .....	<b>294</b>
<b>Appendix B: Consent Forms</b> .....	<b>305</b>
<b>Appendix C: DSM-IV Criteria for Major Depressive Episode</b> .....	<b>314</b>

## LIST OF FIGURES

<i>Number</i>		<i>Page</i>
1.	Mean Yearly Total CES-D Score by K-Cluster .....	158
2.	Scatterplot of MEAN99 by RANGE99 by K-Cluster .....	159
3.	Boxplot of SLOPE99 by K-Cluster .....	160
4.	Scatterplot of SLOPE99 by K-Cluster .....	161
5.	Boxplot of RANGE99 by K-Cluster .....	162
6.	Boxplot of SDCESD by K-Cluster .....	163
7.	Example of K-Cluster 1 Stable Moderate .....	164
8.	Example of K-Cluster 2 Stable Low .....	165
9.	Example of K-Cluster 3 Falling Low .....	166
10.	Example of K-Cluster 4 Rare-1 .....	167
11.	Example of K-Cluster 5 Rare-2 .....	168
12.	Example of K-Cluster 6 Rise&Fall .....	169
13.	Example of K-Cluster 7 Stable Non-Depressed .....	170
14.	Example of K-Cluster 8 Rising Low .....	171
15.	Example of K-Cluster 9 Fall&Rise High .....	172
16.	Algorithm: Options for Parameters in Hierarchical Agglomerative Clustering .....	173
17.	Mean Yearly CES-D Score by Hierarchical Cluster .....	174
18.	Boxplot of RANGE99 by Hierarchical Cluster .....	175
19.	Boxplot of SLOPE99 by Hierarchical Cluster .....	176

20.	<b>Scatterplot of MEAN99 by RANGE99 by Hierarchical Cluster.....</b>	<b>177</b>
21.	<b>Scatterplot of MEAN99 by SLOPE99 by Hierarchical Cluster.....</b>	<b>178</b>
22.	<b>Scatterplot of SLOPE99 by RANGE99 by Hierarchical Cluster.....</b>	<b>179</b>
23.	<b>Minor Seesaw Low No Trend .....</b>	<b>231</b>
24.	<b>Major Seesaw Low No Trend .....</b>	<b>232</b>
25.	<b>Slow Oscillation High No Trend.....</b>	<b>233</b>
26.	<b>Rise &amp; Fall .....</b>	<b>234</b>
27.	<b>Increasing Oscillation Low No Trend .....</b>	<b>235</b>
28.	<b>Decreasing Oscillation High No Trend .....</b>	<b>236</b>
29.	<b>Increasing Oscillation Rising .....</b>	<b>237</b>
30.	<b>Decreasing Oscillation Falling .....</b>	<b>238</b>
31.	<b>Decreasing Oscillation Maximally Falling .....</b>	<b>239</b>
32.	<b>Grand W .....</b>	<b>240</b>
33.	<b>Stable Non-depressed .....</b>	<b>241</b>
34.	<b>Stable Moderate .....</b>	<b>242</b>
35.	<b>Generally Linearly Rising .....</b>	<b>243</b>
36.	<b>Blip Up Late .....</b>	<b>244</b>
37.	<b>Stairstep Up .....</b>	<b>245</b>
38.	<b>Stairstep Down .....</b>	<b>246</b>
39.	<b>Combination: Decreasing Oscillation Falling &amp; Increasing Oscillation .....</b>	<b>247</b>

40.	Scatterplot of MEAN00 by RANGE00.....	248
41.	Scatterplot of MEAN00 by SLOPE00 .....	249
42.	Scatterplot of RANGE00 by SLOPE00 .....	250
43.	Scatterplot of MEAN00 by RANGE00 Reference Lines Added .....	251
44.	Scatterplot of MEAN00 by SLOPE00 Reference Lines Added .....	252
45.	Scatterplot of RANGE00 by SLOPE00 Reference Lines Added .....	253

## LIST OF TABLES

<i>Number</i>		<i>Page</i>
1.	<b>Longitudinal Studies of Depression in Midlife Women: Study Name and Date.....</b>	103
2.	<b>Menopausal Status In Nine Longitudinal Studies Of Depressive Symptoms In Midlife Women .....</b>	104
3.	<b>Longitudinal Studies of Depression during Menopausal Transition: Methods .....</b>	106
4.	<b>Descriptive Statistics: Yearly Total CES-D scores from the Seattle Midlife Women's Health Study .....</b>	124
5.	<b>Iterative or K-Cluster Final Cluster Centers: 9-Cluster Solution; N=216 .....</b>	180
6.	<b>Summary Measures of K-Cluster 9-Solution; N=216.....</b>	181
7.	<b>Summary Measures of Hierarchical Cluster 4-Solution; N=249 .....</b>	182
8.	<b>Descriptive Statistics of Total CES-D Score by Menopausal Transition Stage 1999; N=190.....</b>	183
9.	<b>Yearly ANOVA of Total CES-D Scores by Menopausal Transition Stage .....</b>	184
10.	<b>Yearly Comparison of Total CES-D Scores Between Late Transition and All Other Menopausal Transition Stages Combined.....</b>	185
11.	<b>History of Sexual Abuse/Assault by K-Cluster 9-Solution; N=174 .</b>	186
12.	<b>Historic Trend in Percent Using Prescriptive Antidepressant Drugs .....</b>	187

## **CHAPTER 1: PROBLEM STATEMENT**

Approximately 13% of women in the United States have major depression at any one time (Kessler, McGonagle, Zhao, Nelson, Hughes, Eshleman et al., 1994). Over a lifetime 21.3% of women will experience major depression. Women are twice as likely to experience depression as men (Nolen-Hoeksema, 1990, p. 32; Wolk & Weissman, 1995). Rates of dysthymia, a chronic low-level form of depression, are an additional 3% cross-sectionally and 8% for lifetime occurrence (Kessler et al., 1994). In addition to the personal suffering of the woman with depression, other aspects such as disability, hospitalization and use of health services, and mortality from suicide are problems affecting the individual, women's families, and the rest of society.

In the U.S., women of the "baby boom" generation are now increasingly entering the menopausal transition and postmenopausal phases of their lives. These demographic changes affect both provision of health services and marketing of products such as pharmaceuticals. There are two health problems that have been studied in most research about the menopausal transition and menopause: hot flashes and depression. Many people, lay and professional, assume that women in midlife, approaching and completing menopause, will commonly experience an increase in depression (Coleman, 1993; Schmidt & Rubinow, 1991). Hormone replacement therapy (HRT) has been prescribed for midlife depression, expressing the assumption that the hormonal changes of menopausal transition are related to depression (Burt, Altshuler, & Rasgon, 1998; Schmidt, Nieman, Danaceau, Tobin, Roca, Murphy et al., 2000).

While the diagnosis of involuntional melancholia linked to this assumption is no longer recognized in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, (DSM-IV) (American Psychiatric Association Committee on Nomenclature [APA], 1994), the expectation of increased incidence and prevalence of depression in midlife women remains common (McKinlay, McKinlay, & Brambilla, 1987), despite the lack of epidemiologic evidence for it (Comstock & Helsing, 1976; Craig & Van Natta, 1979; Frerichs, Aneshensel, & Clark, 1981; Gatz & Hurwicz, 1990; Hirschfeld & Cross, 1982; Kessler et al., 1994; Kessler, Borges, & Walters, 1999; Weissman, Bland, Canino, Faravelli, Greenwald, Hwu, et al., 1996). Assumptions about the depressing nature of menopause or the transition to menopause have also not been supported by longitudinal research on community samples of midlife women (Bromberger & Matthews, 1996a, 1996b; Busch, Zonderman, & Costa, Jr., 1994; Hällström & Samuelsson, 1985; Holte, 1992; Kaufert, Gilbert, & Tate, 1992; Woods & Mitchell, 1996).

Nevertheless, clinicians and many women themselves recognize that some midlife women are depressed. Several longitudinal studies on community samples have shown that the single best predictor of depression in midlife is a history of previous depression (Avis, Brambilla, McKinlay, & Vass, 1994; Bromberger & Matthews, 1996b; Busch et al., 1994; Dennerstein, Dudley, & Burger, 1997; Hällström & Samuelsson, 1985; Holte, 1992; Kaufert et al., 1992; Winokur, 1973; Woods & Mitchell, 1996). Concern about the chronic nature of depression and/or its high probability of recurrence, and the concomitant implications for treatment, have begun to appear in the psychological and psychiatric literature (First, Donovan, & Frances, 1996; Keller, Klein, Hirschfeld, Kocsis,

**McCullough, Miller et al., 1995; Keller, Hanks, & Klein, 1996; McCullough, Kornstein, McCullough, Belyea-Caldwell, Kaye, Roberts et al., 1996; Mueller & Leon, 1996).**

Controversies over both longitudinal course and levels of depression are ongoing. For example, the proposed diagnosis of dysthymia requires the persistence of low-level symptoms for at least two years (APA, 1994). Yet for major depressive disorder course modifiers such as “chronic” are optional (Gwirtsman, Blehar, McCullough, Kocsis, & Prien, 1997; Keller et al., 1996). Different levels of depression have been proposed, such as minor or subsyndromal depression (Broadhead, Blazer, & George, 1990; Judd, Rapaport, Paulus, & Brown, 1994; Judd, Akiskal, & Paulus, 1997; Maier, Gansicke, & Weiffenbach, 1997).

In the last decade, the concept of heterogeneity of depression has also been introduced (Judd, 1997a; Judd, 1997b; Winokur, 1997). Double depression refers to a chronic low level of symptoms on which is superimposed a major depressive episode (First et al, 1996; Keller, Hirschfeld, & Hanks, 1997). Additionally, diagnoses such as seasonal affective disorder (SAD) (Betrus & Elmore, 1991; Elmore, Betrus, & Burr, 1994), premenstrual dysphoric disorder (PMDD) (APA, 1994), recurrent brief depression (RBD) (Angst & Hochstrasser, 1994), depressive spectrum disorder or character spectrum disorder (Akiskal, 1988; Winokur, 1997), and depressive personality (Angst & Merikangas, 1997) have continued to fuel discussions.

Further areas of complexity arise from the practice of labeling depression based on its presumed cause, for example endogenous, psychogenic, reactive, or the archaic involuntional depression (Detre, 1968; First et al., 1996; Weissman, 1979). Causal models

**of depression, including those assuming a relationship between menopause and depression, have been hypothesized from several points of view (Avis et al., 1994; McKinlay, McKinlay & Brambilla, 1987):**

- (1) The biological school of thought attributes depression to neurotransmitters, hormones, the menstrual cycle, and/or menopause (Altman, Sachar, Gruen, Halpern, & Eto, 1975; Amsterdam, Winokur, Lucki, & Snyder, 1983; Gruen, Sachar, Altman & Sassin, 1975; Guicheney, Leger, Barrat, Trevoux, Lignièrès, Roques et al., 1988; Markianos, Kykouras, & Stefanis, 1996; Matussek, 1988; Parry, 1989; Post, 1992; Roy, Linnoila, Jimerson, Gold, & Pickar, 1986; Rubinow, Schmidt, & Roca, 1998a, 1998b; Sachar, Altman, Gruen, Glassman, Halpern & Sassin, 1975; Saletu, Brändstatter, Metka, Stamenkovic, Anderer, Semlitsch et al., 1995; Saletu, Brändstatter, Metka, Stamenkovic, Anderer, Semlitsch et al., 1996; Smith & Studd, 1994).**
- (2) The sociocultural school of thought attributes depression to gender roles and social inequality (Kandel, Davies & Raveis, 1985; Nolen-Hoeksema, 1990, Chap 4; Stoppard, 2000, Chap. 5; Woods, 1985), both of which can imply harassment or violence toward girls and women such as child abuse and rape (Stewart & Robinson, 1995). Gender roles have also been proposed as the mechanism whereby menopause might cause depression because of the concomitant loss of fertility, sexual attractiveness, or the empty nest syndrome (Stoppard, 2000, Chap. 9; Wilbur & Dan, 1989).**

- (3) The psychological school of thought proposes that individual personality and/or response styles are causal for depression (Dohrenwend, 1998; Lara & Klein, 1999; Nolen-Hoeksema, 1990, Chap. 6; Skodol, 1998; Stoppard, 2000, Chap. 4). Depression-enhancing response styles have been identified as part of gender socialization for women (Nolen-Hoeksema, 1990, Chap. 7). Child abuse also has been proposed as influencing development of depressive and/or neurotic personality type, which is then associated with adult depression (Johnson, Cohen, Brown, Smailes, & Bernstein, 1999; Widom, 1999).**
- (4) The life events and stressors school of thought attributes depression to a combination of adversity/stress and lack of stress-buffering factors (Brown & Harris, 1978; Dohrenwend, 1998; Moscarello, 1991; Skodol, 1998; Woods & Mitchell, 1997). The biological and/or sociocultural changes of the menopausal transition also can be interpreted to be in this category. Harassment or violence toward girls and women also can be identified as a life event or stressor that is causal for depression (Mackey, Sereika, Weissfeld, Hacker, Zender, & Heard, 1992; Moscarello, 1991; Stewart & Robinson, 1995).**

**It is not surprising that the complexity of the heterogeneity of depression on the one hand and various theories of possible relationships to the menopausal transition on the other has resulted in controversy and some lack of clarity. In order to investigate the heterogeneous aspects of depression and some of the proposed causes of depression in midlife women, an approach that encompasses complexity is required, especially when the proposed cause is itself changing over time such as is the case for menopausal**

**transition. Description of patterns allows such an approach (Crawford, 1982; Nijhout, 1997), the patterns in this case being those of longitudinally repeated measures of depression over time as well as predictors. An investigation of such patterns would allow simultaneously a study of the complexity of depression and predictors of this complexity. With a clearer understanding of the longitudinal patterns and predictors of depression in midlife women, primary, secondary, and tertiary prevention then can be appropriately planned.**

#### **Aims**

**The aims of this study were to:**

- 1. describe and characterize longitudinal patterns of depressive symptoms in midlife women before, during, and after the transition to menopause.**
- 2. explore relationships between longitudinal patterns of depressive symptoms and selected predictors of these patterns or particular changes within patterns.**

**Antidepressant drug use was expected to be a confounder, and the specific predictors were menopausal transition and life events or stressors, including in particular harassment or violence.**

#### **Study Questions**

- 1. Are there women who stay depressed chronically, including during midlife?**
- 2. Are there women who experience short-term depression at different times in midlife?**
- 3. Are there significant relationships between longitudinal patterns of midlife women's depressive symptoms and**
  - a) Menopausal transition?**

**b) Life events and stressors?**

**c) Harassment or violence?**

**The chapters are organized as follows:**

- 1. Problem Statement and Aims.**
- 2. Background and Significance, including definitions, a review of causal models of depression, of studies of the course of depression, and of longitudinal studies of depression in midlife women.**
- 3. Research design and methods from the parent study and general design and methods of the current study applicable to both analyses.**
- 4. Specific methods, results, and focused discussion for a series of cluster analyses, both iterative (or k-cluster) and hierarchical, including review of the analytic methods in case these are not familiar to the reader.**
- 5. Specific methods, results, and focused discussion for a phenomenological analysis, including a review of the analytic method in case this is not familiar to the reader.**
- 6. Discussion and comparison of results from both analyses, including limitations and implications for further research.**

## **CHAPTER 2: BACKGROUND AND SIGNIFICANCE**

**The purpose of this chapter was to review three topics that were background for the current study. A preliminary section reviewing definitions has been followed by a second section briefly reviewing causal models and theories of depression salient for the current study of longitudinal patterns of depressive symptoms in midlife women. The third section is a review of studies of the course of depression in general. The fourth section is an in-depth review of longitudinal studies of depression during the menopausal transition with particular attention to methodological issues.**

### **Definitions**

**In order to study longitudinal patterns of depression in midlife women some terminology surrounding depression and menopause has to be defined. Midlife has been defined as 35-55 years of age (Fogel & Woods, 1995, p. 79). While pattern is a commonly used term, its definition can be a bit elusive. Nevertheless, pattern can be defined as a set of some phenomena with qualities of repeatability or reliability of relationships or processes as central (Crawford, 1982). For longitudinal patterns, the relationships of interest are both from one time point to another and between all time points together as a whole. In considering causes of patterns, however, Nijhout cautions that “a simple pattern gives no clue as to the process that gave rise to it” (1997, p. 8). Depression is a general term (Stoppard, 2000, p. 10). Specific definitions have been presented grouped by three dimensions of depression: severity, duration, and course.**

### Severity

Severity of depression is determined by a simple sum of depressive symptoms (APA, 1994; Chen, Eaton, Gallo, & Nestadt, 2000). Depressive symptoms are the group of symptoms used to diagnose or measure depression. Symptoms include negative affect, positive affect, somatic and retarded activity, and interpersonal subgroups (Radloff, 1977).

### Depressed Mood

Depressed mood is a single symptom (Gath & Iles, 1990). Commonly depressed mood is called just that, “depressed,” but is sometimes referred to as “blue” or “low” or “down in the dumps.” Depressed mood is one of the first two criteria in the DSM-IV (see Appendix C), one of which is required for diagnoses (APA, 1994). Depressed mood is also one of the items included in measures of depression, such as in the Center for Epidemiologic Studies-Depression Scale (CES-D): “I felt depressed” (Radloff, 1977). Some authors have proposed that depressed mood alone is a clinically significant problem worthy of investigation and intervention (Gath & Iles, 1990).

### Depressive Symptoms

Depressive symptoms are a constellation of reported negative experiences identified and grouped as such by professionals (Stoppard, 2000, p. 10). The DSM-IV (APA, 1994) and standardized measures such as the CES-D (Radloff, 1977) contain lists of these symptoms, which all include depressed mood. There is consistency in many of these lists of depressive symptoms, although recently some additional symptoms have been proposed for a new diagnostic category (Keller et al., 1997; Klein, Kocsis,

McCullough, Holzer III, Hirschfeld, & Keller, 1996). Often, negative experiences that are also “associated with” depression are classified as physical and described as somatization (Stewart & Robinson, 1995), reflecting a basic Cartesian dualism underlying modern medical taxonomy.

### Major Depressive Disorder

Major depressive disorder (MDD) is a psychiatric diagnosis “characterized by one or more Major Depressive Episodes” (APA, 1994, p. 317). In an episode, at least five of nine criteria (see Appendix C) must be present for at least two weeks; at least one of the five symptoms must be either 1) depressed mood most of the day, nearly every day or 2) markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (p. 327).

Specifiers, such as single episode or recurrent, may be added to a diagnosis of MDD. A single episode “is considered to have ended when the full criteria for the Major Depressive Episode have not been met for at least 2 consecutive months” (p. 339). In order to qualify as recurrent MDD, a previous episode must have ended (p. 345). A longitudinal course specifier, chronic, may be added to the diagnosis of MDD if “full criteria for a Major Depressive Episode have been met continuously for at least the past 2 years” (p. 382). The DSM-IV diagnosis of MDD, therefore, includes episodes of less than five symptoms if they last less than two months and connect episodes of full MDD.

Within the diagnosis of MDD there is a range of severity from five to nine symptoms. Recently, distinctions between moderate depressive episodes (5-6 symptoms of the DSM-IV criteria) and severe depressive episodes (7-9 symptoms) have been found

in disability, suicide risk, and presumably suffering (Chen et al., 2000; Keller, Lavori, Mueller, Endicott, Coryell, Hirschfeld et al., 1992; Kessler, Zhao, Blazer, & Swartz, 1997). In fact, Chen et al. (2000) found the differences between severe and moderate were similar to those between moderate and mild (or subsyndromal), implying a continuum of a single health problem rather than what Angst & Merikangas (1997) refer to as “threshold” diagnoses.

#### Minor Depressive Disorder And Subsyndromal Depression

The DSM-IV includes a diagnosis of Minor Depressive Disorder (mDD) that entails “episodes of at least 2 weeks of depressive symptoms but with fewer than the five items required for Major Depressive Disorder” (p. 350), that is, requiring depressed mood and at least two other symptoms from the standardized list of criteria (see Appendix C). Broadhead et al. in 1990 were among the first to propose the label subsyndromal depression to describe depressed mood and at least one other symptom. There is, therefore, an overlap with mDD when subsyndromal depression shows at least two other symptoms in addition to depressed mood. Subsyndromal and minor depression have been recognized as clinically significant in terms of disability (Broadhead et al., 1990; Gotlib, Lewinsohn, & Seeley, 1995; Johnson, Weissman, & Klerman, 1992; Judd et al., 1994).

#### Dysthymic Disorder Or Dysthymia

Dysthymic disorder is a diagnosis based on chronically depressed mood for at least two years accompanied by at least two other of the criteria for MDD excluding suicidality, although some additional symptoms have been proposed for the standardized

**list of dysthymia symptoms (Keller et al., 1996; Keller et al., 1997; Klein et al., 1996). Because dysthymic disorder could be seen as mDD with the course modifier required, there is some controversy regarding its definition (Brieger & Marneros, 1997; Gwirtsman et al., 1997; Keller et al., 1996; Keller et al., 1995; Maier et al., 1997; Shea & Hirschfeld, 1996).**

**In sum, depressed mood is a single-dimension phenomenon, in lay terms often referred to as feeling depressed or blue. Subsyndromal depression is depressed mood plus at least one other symptom from the standardized list of criteria (see Appendix C). Minor depressive disorder is depressed mood plus two to four other symptoms from the standardized list of criteria, lasting every day for at least two weeks. MDD is depressed mood (or loss of pleasure) plus at least four others from the same standardized list, again lasting every day for at least two weeks. These four labels therefore exist on a continuum of increasing numbers of symptoms, with an overlap between subsyndromal and mDD (Judd, 1997b). In addition, a duration of the disorders is required (at least two weeks). The disorders are also defined by an episode based on subsequent relief from the symptoms for another specified duration (at least two months).**

#### **Duration**

**As early as 1976, Craig and van Natta proposed that duration of symptoms was crucial for diagnosis. Recurrent brief depression (RBD) is a diagnosis requiring the same severity of symptoms as MDD but lasting less than two weeks, thereby not meeting the duration criteria of MDD while still causing dysfunction and disability (Angst & Hochstrasser, 1994; APA, 1994; Lepine, Pelissolo, Weiller, Boyer, and Lecrubier, 1995).**

**On average, an episode of RBD lasts two to three days. RBD also requires a course designation, so that it must recur at least 10-12 times per year (see next section).**

**Prevalence of RBD was approximately 16% in the general population in Switzerland, so it is of clinical concern. Dysthymia and RBD were proposed because the two-week duration specified in mDD and MDD was too rigidly limiting in defining clinical phenomena. The implication that duration of episode is an important part of diagnosis, not necessarily an optional modifier, has been addressed in the two diagnoses of RBD and dysthymia. In sum, RBD, unspecified MDD, and MDD with a chronic specifier might be seen as existing on a continuum of increasing duration of episode. Similarly, mDD and dysthymia can be seen as existing on a continuum of increasing duration, although at a lower level of severity than MDD.**

#### **Course**

**Course can be defined as longitudinal pattern of diagnoses, where longitudinal pattern can be seen as a broader term than course, not limited to diagnostic criteria. Heterogeneity in longitudinal pattern has been addressed by requiring or allowing specification in diagnosis of duration of episode, duration of post-episode or interval, and frequency of recurrence. Full remission is defined as the complete relief of depressive symptoms lasting for at least two months after a major depressive episode. Recovery is applied to the period of complete relief of symptoms following full remission. Partial remission is relief of only some depressive symptoms after a major depressive episode but where the number of symptoms no longer meets MDD diagnostic criteria. Relapse is defined as a subsequent major depressive episode after full remission and implies the**

**chronic nature of MDD. Recurrence is a subsequent major depressive episode after a recovery and also implies chronicity of the disorder (Bland, 1997; Mueller & Leon, 1996).**

**Double depression has been proposed as a diagnosis of MDD that follows dysthymia, or is superimposed on it (Akiskal, 1994; First et al., 1996; Friedman & Kocsis, 1996; Keitner, Ryan, Miller, Kohn, & Epstein, 1991; Keller et al., 1995; Keller et al., 1997; Kessler et al., 1997; Miller & Keitner, 1996). Since dysthymia proceeds to MDD in the large majority of cases, the diagnosis of dysthymia has been interpreted alternatively as a prodrome and/or partial remission of a major depressive episode (Horwath, Johnson, Klerman, & Weissman, 1992; Keller et al., 1995). Double depression has been identified, however, as one of the most disabling forms of depression, thus justifying the recognition in a diagnosis (Keller et al., 1995; Keller et al., 1997). Dysthymia, double depression, RBD, and the course specifiers chronic or recurrent include, therefore, the sense of longitudinal pattern.**

**Attempts to combine simultaneously aspects of course such as duration, recurrence, and severity in diagnosis have not been entirely successful, however (First et al., 1996). The diagnoses of RBD and dysthymia are cases in point. RBD meets all the requirements of MDD except that the duration of symptoms for at least two weeks is not met (Angst & Hochstrasser, 1994; Lepine et al., 1995). However, this diagnosis also requires a course specification of at least 10-12 episodes per year, as well as the duration specification. Dysthymia is the only proposed diagnosis (APA, 1994) that requires a chronic duration (two years of unrelieved symptoms). In contrast, two years of**

unrelieved symptoms of *major* depressive disorder would simply have an added optional course modifier of chronic. Yet dysthymia also requires fewer numbers of symptoms than MDD, implying it includes an indication of severity. Minor depressive disorder and/or subsyndromal depressive disorder are also less severe than major, with fewer symptoms required, but no longitudinal course specified. In addition, according to some authors, dysthymia encompasses somewhat different, although for the most part overlapping, symptoms as MDD (Keller et al., 1995).

In sum, both severity and duration of episode, plus duration of post-episode, and occasionally longitudinal course, have been included, albeit inconsistently, in diagnoses (First et al., 1996; McCullough et al., 1996). The attempts over the years to differentiate depressions in terms of only some of the dimensions of severity, duration, post-episode duration, and course have led to this rather inconsistent proliferation of labels and diagnoses.

### Causal Models Of Depression

What has been labeled the heterogeneity of depression has been discussed within the past 10 or so years by researchers proposing that depression encompasses a spectrum of diagnoses that reflect cause in addition to severity, duration, and course (Akiskal, 1985; Chen et al., 2000; Judd et al., 1994; Keller et al., 1995; Keller et al., 1997; Winokur, 1997). Causal labels such as endogenous, reactive, involuntal, vascular, seasonal affective, character spectrum, and premenstrual dysphoric have all been used.

Causal attributions of depression can be grouped into several categories including, biological, socio-cultural, psychological, and life events. Proposed biological causes of

depression are the basis for the rationale to include measurement of depression in virtually all studies of the menopausal transition. According to these theories, hormonal changes during the menopausal transition or at menopause could be assumed to cause depression in women, either directly, or indirectly through causing symptoms such as hot flashes, or even more indirectly through hot flashes that cause insomnia which in turn causes depression via sleep deprivation (Avis et al., 1994; Baker, Simpson, & Dawson, 1997; Thomson, 1976). Biological causal theories are often based in the epidemiologic observation that rates of depression are approximately twice as high in women as in men in the U.S. (Bracke, 1998; Breslau, Schultz, & Peterson, 1995; Briscoe, 1982; Gjerde, Block, & Block, 1988; Hankin, Abramson, Moffitt, Silva, McGee, & Agnell, 1998; Jenkins, 1985; Nolen-Hoeksema, 1990, Chap. 2; Ross & Mirowsky, 1984), with the unstated assumption that the sex differences in depression are due to physiological differences. Nevertheless, biological and especially hormonal causes of depression have been studied extensively with virtually no reliable evidence to support these theories (Nolen-Hoeksema, 1994, Chap. 3; Pardo et al., 1993; Rubinow, Hoban, Grover, Galloway, Roy-Byrne, Andersen et al., 1988; Ussher, 1992).

The socio-cultural causal theories attribute depressive symptoms in women to oppressive gender roles and social inequality (Kandel et al., 1985; Lara & Klein, 1999; Nolen-Hoeksema, 1990, Chap. 4; Pearlin, 1975; Wolk & Weissman, 1995; Woods, 1985). Socio-cultural aspects of menopause, such as loss of fertility and sexual attractiveness, and empty nest syndrome (Lennon, 1987; Stoppard, 2000, Chap. 9), have been proposed as causes of depression in midlife. There can be some overlap with the

**school of thought that considers life events and stressors, since part of the oppression of women can be life events and/or stressors that are effects of the poverty, harassment, and violence of social inequality. According to Schafran's review, "a major reason women have higher rates of depression than men is the high level of rape, battering, sexual harassment, and child sexual abuse in their lives" (1996, p. 16). Several researchers have found relationships between childhood abuse or adult rape and adult depression (Goldberg, 1994; Hall, Sachs, Rayens, & Lutenbacher, 1993; Kilpatrick, Resnick, Saunders, Best, 1998; Liem, James, O'Toole, & Boudewyn, 1997; Mackey et al., 1992; Scholle, Rost, & Golding, 1998; Moscarello, 1991; Stewart & Robinson, 1995; Walker, Katon, Hansom, Harrop-Griffiths, Holm, Jones et al., 1992; Widom, 1998).**

**Psychological theories posit stable characteristics of the individual such as personality type and/or response styles as most predictive of depression (Akiskal, 1985; Akiskal, 1988; Akiskal, Hirschfeld, & Yerevanian, 1983; Bolger & Schilling, 1991; Carver & Scheier, 1990; Costa, Jr., Zonderman, McCrae, Comoni-Huntley, Locke, & Barbano, 1987; Duncan-Jones, Fergusson, Ormel, & Horwood, 1990; Hirschfeld, Klerman, Clayton, & Keller, 1983; Lara & Klein, 1999; Nolen-Hoeksema, 1990, Chap. 5, 6; Ormel, Stewart, & Sanderman, 1989; Ormel & Wohlfarth, 1991; Rohde, Lewinsohn, Tilson, & Seeley, 1990; Tiggemann, Winefeld, Winefeld, & Goldney, 1991; Watson & Clark, 1984; Wolk & Weissman, 1995; Young, Fogg, Scheftner, Fawcett, Akiskal, & Maser, 1996). Again, there can be overlap with other causal models since some authors have proposed that childhood traumatic life events, such as childhood sexual abuse, can determine or influence personality type (Johnson, et al., 1999; Walker & Katon, 1996;**

Widom, 1998). Nolen-Hoeksema has proposed that gender is related to response style, which in turn enhances the development of MDD (1990, Chap. 7).

Other researchers hypothesize interactions between life events, stressors, and resources or buffers (Ensel & Lin, 1991; Nuckolls, Cassel, & Kaplan, 1972; Silberg, Heath, Kessler, Neale, Meyer, Eaves, & Kendler, 1990; Woods & Mitchell, 1996; Woods & Mitchell, 1997). Once again, menopause has been interpreted as a life event and/or stressor that could cause depression. Harassment, violence, and effects of poverty could also be included under this rubric (Mackey et al., 1992; Moscarello, 1991; Stewart & Robinson, 1995).

Only a few authors have noted that schools of thought (and research) have developed for each area of causal theory, apparently independently, without much inter-communication taking place (Lerner, 1987). For example, few causal models of depression in the psychological literature mention menopause or indeed any biological factors as causal or even predictive of depression. On the other hand, biologic-focused researchers rarely recognize social, cultural, or psychological factors in causal models (Katschnig, Angst, Clayton, Gershon, Hautsinger, Helmchen et al., 1983; Lerner, 1987), even though social and psychological, not biological, aspects of menopause might well be causal for depression.

In recent years causal models of depression have been proposed (Ormel & Wohlfarth, 1991; Walker & Gelfand, 1997) that have attempted to interrelate a fairly complex set of variables. A stable factor, such as personality type is proposed as an important determinant of depression or its absence. Personality type is generally

accepted as established before adulthood (Nolen-Hoeksema, 1990; Chap. 5), and it might possibly be determined or influenced by childhood life events. Some factor to explain adult depressive episodes, such as life events and/or long-term stressors, then is proposed to explain depressive episodes and especially changes in episodes, such as persistence, remission, recovery, relapse, or recurrences. Presumably the menopausal transition could be such a stressor. However, in endogenous depression, by definition, it is assumed there are no specific triggers for a depressive episode. At the other end, Duncan-Jones et al. (1990) also asserted that researchers should not overlook or take for granted factors that cause or influence stability. Lastly, although not many of the cited authors included it, some factor to explain changes for the minority of women whose general level of depression and/or pattern of fluctuation over time increases or decreases (see Chapter 5) would also have to be included, such as for women whose first major depressive episode occurs during midlife and the menopausal transition.

Thus according to a complex theory, depressive symptoms in general (or their lack) may be seen as dependent upon two factors that interact: stable characteristics of the individual and modifiers or triggers. Evidence for a stable characteristic predicting depression has been fairly consistent. Ormel & Wohlfarth (1991) found in their multidimensional causal modeling that personality type, labeled neuroticism, was by far the major contributor to depression at both of two time points. Neuroticism is a personality type characterized by negative emotionality and negative temperament (Skodol, 1998), and additionally “a tendency to have unrealistic ideas, an inability to control urges, and inefficient ways of coping with stress” according to Ormel and

**Wohlfarth (1991, p. 744). This personality type is commonly operationalized as complaining of many symptoms (Duncan-Jones et al., 1990). These symptoms can include those stereotypically associated with the menopausal transition. However, Ormel and Wohlfarth also discussed that if personality causes depression, the study question then becomes what causes that personality type: “The reason why [some people] have a tendency to experience distress [i.e., depression] remains an open question for future research to answer. Similarly, we do not know how this enduring disposition to become distressed comes about” (p. 752).**

**In the complex model of depression, personality and life events then are proposed to interact in adulthood. Certain personality types are more likely to be “risk takers” or more likely to put themselves in harm’s way. This same personality type also makes people less able to extricate themselves from the risk or recover from the harm. In general, behaviors that reflect the personality, such as inappropriate interpretation of threatening cues from the environment and/or difficulties with social skills (Walker, Gelfand, Katon, Koss, Von Korff, Bernstein et al., 1999), lead to increased number and severity of negative life events (Fergusson & Horwood, 1987). For example, survivors of childhood sexual abuse are more likely to be raped as adults (Kilpatrick et al., 1998; Mackey et al., 1992).**

**Then, too, in the complex model, life events can be experienced and interpreted by the individual in different ways. For women with depressive personalities, adversity or difficult life events might provoke the response of a depressive episode (Lara & Klein, 1999; Wolk & Weissman, 1995). For some women the same adversity or difficulty**

might be experienced as a challenge. The latter women's healthy outcome is characterized as resilience (Liem et al., 1997). Nevertheless, some life events might be so catastrophic that they overwhelm even healthy women. For example, overwhelmingly severe or numbers of negative life events, such as death of a child, or loss of several family members and a house in the same earthquake (Giel, 1998), or adult rape (Mackey et al., 1992; Moscarello, 1991; Schafran, 1996; Stewart & Robinson, 1995) lead to depression even for those women whose personality does not predispose them to risk-taking or depression.

Factors that modify personality development and interaction with life events have not been well studied. Life events during childhood, especially loss of a mother (Brown & Harris, 1984, Chap. 1; Lara & Klein, 1999; Wolk & Weissman, 1995) and environmental disasters such as earthquakes, or trauma such as caused by wars (Giel, 1998; Mollica, Poole, & Tor, 1998) have been proposed as affecting the stable characteristic that has been labeled personality. A few researchers have proposed that childhood abuse is a major factor in creating a depressive and/or neurotic personality type (Hall et al., 1993; Johnson, et al., 1999; Walker et al., 1997; Walker et al., 1999; Widom, 1998). The childhood trauma can be individual, such as abuse, or commonly shared, such as devastation of a whole village from an earthquake, without differing effects (Dohrenwend, 1998).

As for factors affecting stability and lability, Duncan-Jones et al. (1990) proposed that in women who have recovered from a major depressive episode, the stabilizing effect of continuing psychotherapy, for example, should be included in any causal model. Liem

et al. (1997) did investigate factors in adults that predicted resilience (i.e., absence of depression) after childhood sexual abuse. Not surprisingly, longer amount of time since the abuse, male gender, and a childhood sense of self-efficacy (perhaps not unrelated to male gender) were predictors of absence of depression in adulthood. Although no definitive explanation was discussed, large number of siblings also predicted absence of depression in adulthood after a history of childhood sexual abuse.

As Lara and Klein (1999), Nolen-Hoeksema (1990), and Wolk and Weissman (1995) have noted, many causal models of depression are lacking in empirical bases. The few research studies attempting to probe causal pathways have been inconclusive for methodological reasons and/or are conflicting in results. In particular the role of life events and stressors in predicting depression has been inconsistent. Chen et al. (2000) and Ormel and Wohlfarth (1991) found life events to contribute little to prediction of depression severity. On the other hand, Tiggeman et al. (1991) found life events significant.

One primary methodological problem is the retrospective ascertainment of both depressive episodes and especially of the supposed causes such as life events. Retrospective measurement in and of itself is subject to the transformation memory imposes. It has also been recognized that current mood affects memory (Mundt, Reck, Backenstrass, Kronmüller, & Fiedler, 2000). Recall of events including past depressive episodes, even within the past three months, let alone past six or 12 months or longer, is likely to be affected by current mood at time of recall.

A second basic methodological problem is measurement of various confounders and predictors by self-report. Current mood also affects measurement of concurrent variables, such as social support or other presumed buffers, health status, income adequacy, life events, supportiveness of partner, and stress (Aldwin, Levenson, Spiro III, & Bossé, 1989; Duncan-Jones et al., 1990; Henderson, 1983; Henderson, 1998; Rohde, Lewinsohn, & Seeley, 1990; Rosenbaum, Lewinsohn, & Gotlib, 1996; Tiggemann et al., 1991). The concurrent mood is itself a confounder, while at the same time it is often the outcome variable. Complex modeling such as with structural equations would be required to investigate multiple and multi-directional relationships (Byrne, 1989). When concerns about the effect of current mood on the self-report of these variables is combined with concerns about measurement of them by retrospective recall, the interpretation of studies employing these methods becomes problematic.

Thirdly, many studies of depression do not include gender as a factor. Depression in men and in women are often considered equivalent. As the feminist postmodernist Stoppard (2000, Chap. 1) pointed out, however, it might not be possible to consider depression without including political, social, cultural, and economic influences of gender on women's lives. There are a few empiric-analytic studies that have identified gender differences in the depression itself, not just in predictors. In general, women were more likely to express their emotions verbally, more likely to somatize, more likely to experience an increase in negative affect rather than a decrease in positive as men do, and to withdraw socially rather than become aggressive and violent (Bracke, 1998; Briscoe,

1982; Gjerde et al., 1988; Jenkins, 1985; Liem et al., 1997; Widom, 1998; Williams, Spitzer, Linzer, Kroenke, Hahn, DeGruy et al., 1995).

In summary, causal models of depression have fairly consistently identified that a stable characteristic of the individual, usually labeled personality type, is a good predictor of adult depressive disorder, but life events or other proposed predictors have not been reliably identified to predict depressive episodes. Biological causes have also not been reliably identified. Methodologically, causal attribution would require prospective longitudinal research methods, adjusted for current mood of all covariates and predictors including gender, and use community sampling that included institutionalized adults.

#### Studies of the Course Of Depression

As has been noted, diagnoses of depression and concepts about cause are based on an understanding of course or longitudinal patterns of depression (Angst, 1988). Course is of particular interest because aspects of it have important prognostic value and implications for tertiary prevention. A person who has had two major depressive episodes has a 90% chance of having a third; dysthymia in 80% of cases leads to and/or follows a major depressive episode (Gonzales, Lewinsohn, & Clarke, 1985; Judd et al. 1997; Keller et al., 1996). In addition, effectiveness of treatments have been noted to vary depending on different diagnoses that reflect causal attribution (Akiskal, 1994).

Two approaches have been taken to studying the course of depression. One could be termed the group approach. In this case, the percentage of people still depressed, remitted, recovered, and with recurrence of a major depressive episode is measured at particular intervals, often using survival analysis-type techniques. The approach may be

described as a repeated cross-sectional measurement at various time intervals (Evans, Hollon, DeRubeis, Piasecki, Grove, Garvey et al., 1992; Giles, Jarrett, Biggs, Guzick, & Rush, 1989; Gonzales et al., 1985; Huba, Lawlor, Stallone, & Fieve, 1976; Keitner et al., 1991; Keller, Lavori, Lewis, & Klerman, 1983; Keller et al., 1992; Keller, Lavori, Rice, Coryell, & Hirschfeld, 1986; Kupfer, Frank, Perel, Cornes, Mallinger, Thase, et al., 1992; Lavori, Dawson & Mueller, 1994; Lewinsohn, Zeiss, & Duncan, 1989; Mueller & Leon, 1996; Mundt et al., 2000; Peselow, Dunner, Fieve, & Difiglia, 1991; Shea, Elkin, Imber, Sotsky, Watkins, Collins et al., 1992; Simon, 2000; Warner, Weissman, Fendrich, Wickramaratne, & Moreau, 1992; Wells, Burnam, Rogers, Hays, & Camp, 1992). A second approach to studying course of depression could be labeled individual. In this approach various patterns of depressive symptoms, including severity and duration, are pictured from each of the sampled individuals (Akiskal, King, Rosenthal, Robinson, & Scott-Strauss, 1981; Akiskal, 1982; Angst & Merikangas, 1997; Bracke, 1998; Callahan, Hui, Nienaber, Musick, & Tierney, 1994; Coryell, 1997; Coryell, Endicott, & Keller, 1991; Coryell, Scheftner, Keller, Endicott, Maser, & Klerman, 1993; Costa, Jr. et al., 1987; Furukawa, Konno, Morinobu, Harai, Kitamura, & Takahashi, 2000; Gaynes, Magruder, Burns, Wagner, Yarnall, & Broadhead, 1999; Harlow, Goldberg, & Comstock, 1991; Keller et al., 1995; Kessler et al., 1997; Lewinsohn, Fenn, Stanton, & Franklin, 1986; Lin, Katon, Von Korff, Russo, Simon, Bush, et al., 1998; Matussek, 1988; Turner & Noh, 1988; Winokur, 1997; Zonderman, Herbst, Schmidt, Costa, Jr., & McCrae, 1993).

Prospective research has been limited often to clinical samples starting with diagnosis of depression or at least depressive symptoms and then following the course, usually only until the next episode or a fixed end point in time, whichever comes first. (All of the studies listed in the previous paragraph could be described in this manner except those listed as sampling from community populations. See below.) Recent commentary from a feminist and/or social constructionist point of view has criticized the focus in much depression research on only those women who enter the health care system and receive diagnoses (Stoppard, 2000, Chap. 1). Some studies have been able to sample from general community populations (Angst & Merikangas, 1997; Bracke, 1998; Chen et al., 2000; Costa, Jr. et al., 1987; Harlow et al., 1991; Horwath et al., 1992; Judd et al., 1997; Kessler et al., 1997; Lewinsohn et al., 1986; Lewinsohn et al., 1989; Maier et al., 1997; Turner & Noh, 1988). Coryell et al. (1991) and Judd et al. (1997) reported follow-up of the portion of samples that was without depressive symptoms at the first measurement point, although they both used sampling of family members of depressed patients. Prospective investigation into lifelong course is rare, although Angst and Merikangas (1997) have published results from a 15-year longitudinal study conducted in Zurich, with a measurement interval of every three years. However, the majority of prospective longitudinal studies follow patients no more than two years, with the exception of Bracke et al. (1998) three years; Coryell (1997) 10 years; Coryell et al. (1991) six years; Coryell et al. (1993) five years; Costa, Jr. et al. (1987) nine years; Keller et al. (1992) five years; Keller et al. (1983) five years; Kupfer et al., 1992 three years; Lavori et al. (1994) five years; Peselow et al. (1991) five years; and Turner and Noh

(1988) four years. Retrospective research has been more common, especially in terms of determining course of symptoms and/or disorders by recall. Occasionally the concept of course is reduced to recall of age at onset and number of lifetime episodes (Chen et al., 2000).

Researchers have described a consistent picture of course of depression in women (Angst & Merikangas, 1997; Bracke, 1998; Chen et al., 2000; Keller et al., 1995). A marked increase in incidence rate of MDD occurs in late adolescence (Gjerde et al., 1988; Hankin et al., 1998; Nolen-Hoeksema, 1990, Chap. 3), when prevalence rates for girls first increase above rates for boys. Incidence rates in women after adolescence have been little studied; Hällström & Samuelsson (1985) did measure incidences in midlife women and found them unchanging across the menopausal transition, albeit through retrospective recall of depressive episodes. Prevalence rates are highest for women in adulthood (Kessler et al., 1994) and show slight decreases or steady rates in older years (Ballinger, 1990; Busch et al., 1994; McKinlay, Brambilla, & Posner, 1992).

Along these same lines of approach, cross-sectional studies of the epidemiology of depressive disorders has been repeated from decade to decade (Comstock & Helsing, 1976; Craig & Van Natta, 1979; Frerichs et al., 1981; Hirschfeld & Cross, 1982; Kessler et al., 1999; Kessler et al., 1994; Weissman et al., 1996). If cohort effects are taken into account, these studies provide prevalences at different ages and give some clues as to longitudinal course. Consistently, there is no increase in prevalence of depressive disorders during midlife.

It was noticed early by clinicians (Akiskal, 1982) that many depressed patients were repeaters. Yet, general acceptance and articulation of depression as a chronic disease did not emerge until the last decade or so (Costa, Jr. et al., 1987; Horwath et al., 1992; Keller et al., 1995). In the large community-based study by Keller and colleagues (1995) the most common longitudinal patterns of depressive disorders were 1) “recurrent, with antecedent dysthymia, without full inter-episode recovery”, 2) “single episode without antecedent dysthymia,” 3) “recurrent, without antecedent dysthymia, with full inter-episode recovery,” and 4) “recurrent, without antecedent dysthymia, without full inter-episode recovery” (p. 845). Three of the four most common patterns included recurrence. Risk factors for longer duration, shorter interval, and higher number of recurrences or relapses have also been studied. In general, lack of treatment, a preceding dysthymia or low level of symptoms, remittance without full recovery, a history of previous major depressive episodes, and female gender have been implicated (Akiskal, 1982; Akiskal, 1994; Bracke, 1998; Chen et al., 2000; Coryell et al., 1991; Coryell et al., 1993; Coryell, 1997; Costa, Jr. et al., 1987; Detre, 1968; Furukawa et al., 2000; Gaynes et al., 1999; Giles et al., 1989; Gonzales et al., 1985; Harlow et al., 1991; Horwath et al., 1992; Judd, 1997b; Judd et al., 1997; Keller et al., 1983; Keller et al., 1986; Kessler et al., 1997; Lavori et al., 1994; Lewinsohn et al., 1986; Lewinsohn et al., 1989; Lin et al., 1998; Maier et al., 1997; Matussek, 1988; Mundt et al., 2000; Parry, 1989; Peselow et al., 1991; Simon, 2000; Turner & Noh, 1988; Warner et al., 1992; Winokur 1997). In a longitudinal study Judd et al. (1997) found both persistence of symptoms and surprising lability in number of symptoms, including 22% of people with no symptoms at first

measurement reporting symptoms one year later. Thus, both chronicity and lability were identified.

A comprehensive approach to observing the whole spectrum of longitudinal patterns of depressive symptoms, before carving out single complex patterns for a diagnostic label, has been missing. Hodgepodge single-focus approaches to developing diagnoses, such as seen in proponents of dysthymia and “double” depression, have ignored the entirety of the phenomena. However, the proliferation in the last decade of various diagnoses as well as course and severity modifiers can be seen as an attempt to acknowledge the complexities of the heterogeneity of depression, especially with regard to severity, duration, and course.

Most research has been limited by concepts of diagnoses and thresholds, such as depressed/non-depressed, although there have been a few studies that investigated more complex concepts such as levels of severity (severe, moderate, mild, none) or simply numbers of symptoms (Angst & Merikangas, 1997; Chen et al., 2000; Judd et al., 1997). Researchers have not yet conducted a prospective longitudinal study that starts with a clearly articulated and measured set of dimensions without conflating them. Such a study would 1) remove threshold definitions of depression 2) remove limits of clinical sampling 3) measure over long enough time and frequently enough to identify patterns adequately 4) use research designs and methods that are free from the limits of thinking of course of disease.

### Longitudinal Studies Of Depression And Menopause

An extensive body of research has investigated depression and menopause (Greene, 1992; Nicol-Smith, 1996). Greene reviewed 25 studies and proposed that only those women with the same risk factors for depression at any time of life (low socioeconomic status (SES), stressful marriage, unfulfilling employment, bereavement) were vulnerable to depression, and then more likely during the perimenopause than at menopause itself. Nicol-Smith (1996) reviewed 94 studies and concluded that the results were not consistent enough to determine a relationship between depression and menopause. In fact, only a few of the studies that she presented graphically demonstrated any relationship, and those few were inconsistent in when or which stages of menopausal transition or postmenopause there might be a relationship to elevated depressive symptoms. While she did not critique methods specifically, most of the research has been cross-sectional and often sampled from clinic populations. Research using these methods is still being conducted (Sagsoz, 2001).

As Nicol-Smith reviewed, correlation does not imply causality (Kleinbaum, Kupper, and Muller, 1988, p. 39; Rothman, 1986, Chap. 2). Attempts to link menopause causally to depression would have to consider a temporal relationship, requiring studies with longitudinal design. Nine longitudinal studies of midlife women were identified in the literature, each of which measured depressive symptoms or mental health diagnoses in the same women at more than one time point during the transition to and after menopause. To date, only one longitudinal analysis for each study has been published, with the exception of Bromberger & Matthews (1996a, 1996b) and Hunter (1990, 1992),

who have published twice. Neither second article had additional information about depressive symptoms and menopausal status. Table 1 has provided a summary of study names and researchers. They are arranged in order from smallest to largest geographical area in the U.S., other English-speaking countries, and lastly other countries.<sup>1</sup>

Several of the studies reviewed here also have published results from cross-sectional analyses based on data obtained at entry to the longitudinal study or on earlier or pilot versions of the same study. It could also be expected that further analyses and results from these longitudinal studies will follow in the future. Specifically, the SMWHS has measured hormonal assays and developed clarified and precise categories of menopausal and menstrual status (Woods, Mitchell, & Mariella, 2000) that have been included in analyses published at later dates than the longitudinal study of depressed mood.

The following review was limited to the published, longitudinal portion of these nine studies of depressive symptoms during the menopausal transition. Only the outcome variable of depressive symptoms was reviewed here. Where both cross-sectional and

---

<sup>1</sup> The Study of Women's Health Across the Nation or SWAN study, S. McKinlay, P. I. a multi-site, multi-ethnic longitudinal study of the menopausal transition based in Boston, is also being conducted currently; however, these researchers have not finished data collection and have not published a longitudinal analysis as of this literature review (Sowers, Crawford, Sternfeld, Morganstein, Gold, Greendale et al., 2000). A longitudinal pilot project, using a convenience sample of 11 women and published as a conference poster session, did not specifically mention depression or mental health symptoms, and is not included in this review (Bareford, 1990). Lastly, Collins & Landgren (1997) studied negative mood and psychological symptoms in midlife women published in a paper with the words "longitudinal follow-up" in the title; however, it was a cross-sectional analysis. These authors have not published a longitudinal analysis to date.

longitudinal analyses were included in the same paper, information about the longitudinal portion alone was extracted for this review. Anecdotal information about methods, while available to this author particularly for the SMWHS have not been included for consideration except where cited as personal communications. However, all details of methods often were not available in the single article that published the longitudinal analysis. Cross-sectional, pilot phase, and other publications from the same research project were used as needed to glean the full details of methods.

This review of methodological issues and design focused on sampling methods, size and eligibility criteria, socio-cultural characteristics of samples, cohort effects, measurement tools of depression, definitions of menopausal and menstrual status, other independent variables included, study designs, and analysis strategies. A summary of these elements from each study has been provided in Tables 2 and 3. Lastly, a review of results in light of methods, design, and analytic strategy completed the purpose of this literature review.

## Samples

### Sampling Methods

Most of the longitudinal studies reviewed here used population- or community-based sampling frames such as tax or community registers (Hällström & Samuelsson, 1985; Holte, 1992), representative stratified probability sampling (Busch et al., 1994), random or stratified sampling of total populations from census lists (Avis et al., 1994), from telephone numbers (Dennerstein et al., 1997; Woods & Mitchell, 1996), and from driver's licenses (Bromberger & Matthews, 1996b). Because of the large number of

women available in Gothenburg, Hällström and Samuelsson sampled by day of birth within year of birth (Samuelsson, 1982). All sampling methods of the longitudinal studies produced high participation rates of those contacted and/or eligible for entry to the cross-sectional portions of the studies.

In their longitudinal analysis article, Kaufert et al. (1992) described their sampling as “not truly random” (p. 154). In the more detailed presentation of their methods in an earlier article, however, these authors described a “stratified random sample of the general population of women in this age group [40-59] whose names appeared in either the Henderson’s Directories or the Voters Registration Lists for the rural areas” living in Manitoba province, Canada (Kaufert, 1984, p. 280). No further explanation was given explaining the nature of Henderson’s Directories or the change in description from “stratified random” to “not truly random.”

In spite of criticism by the Norwegian and Swedish authors in particular, the Seattle, University of Pittsburgh, and Melbourne studies justified their sampling frames as representative due to the near universality of telephones and driver’s licenses (Anderson, Nelson, & Wilson, 1998). However, even if the various sampling frames were comprehensive and comparable, response rates varied as well. The Seattle study enrolled 62% (508 of 820) of those eligible women who were contacted by telephone, although true response rates were slightly higher since enrollment of eligible women had to be terminated due to lack of resources, according to the Principal Investigator (Nancy F. Woods, personal communication). The University of Pittsburgh study found 60% of eligible, that is healthy and premenopausal, women volunteered (Matthews, Wing,

Kuller, Meilahn, & Kelsey, 1990). The Melbourne study's authors (Dennerstein, Smith, Morse, Burger, Green, Hopper et al., 1993) explained first that 13% of phone numbers were not included in their sampling frame (these telephone numbers were unlisted in the directory used) and response rate among those eligible and contactable was 70.6%.

While all these response rates were reported as good to excellent, they were lower than that of the Gothenburg study of 89% (800 of 899) at entry to their first wave. The differences might have produced samples that were not entirely comparable. Response rates for other studies fell between the low of 60% in Pittsburgh and the high of 89% in Gothenburg.

Hunter was the exception to community-based sampling, having recruited "volunteers attending an ovarian screening programme at King's College Hospital...recruited through newspaper advertisements and a radio programme" (1986, p. 218). This setting was described as a "non-menopause clinic;" however, the phrase "ovarian screening" is not used in the U. S. and was not further explained by Hunter. Unfortunately, in her discussion Hunter dismissed the importance of the clinical and volunteer nature of her sample when comparing her results to Holte: "the discrepancy between these findings cannot be easily explained by obvious problems of methodology" (Hunter, 1990, p. 363). Other studies have, indeed, explained discrepancies in findings because of the differences between clinical or volunteer and community sampling (Kaufert & Syrotuik, 1981; McKinlay et al., 1987b; Morse, Smith, Dennerstein, Green, Hopper, & Burger, 1994). Consistent with this expected sampling bias, Hunter found

that 51% of her sample had been depressed recently before entry to the study, a prevalence rate at least twice as high as other studies reviewed here.

### Sample Sizes

Sample sizes for the longitudinal portions of the studies ranged from a low of 36 (Hunter, 1990) to a high of 2352 (Avis et al., 1994). These sample sizes reflected not only the original geographic areas, ranging from selected census tracts within city limits (Woods & Mitchell, 1996) to the entire U. S. (Busch et al., 1994), but also the exclusion criteria for the longitudinal portion of studies. The exclusion criteria often eliminated large groups, even the majority, of the original cross-sectional sample (Busch et al., 1994; Holte, 1992; Hunter, 1990; Kaufert et al., 1992).

In the National Health Examination Follow-Up Study (NHEFS) (Busch et al., 1994) the total sample size used for longitudinal analysis of depressive symptoms was never presented as such in the published article. Different measures of psychological distress were administered at Time 1 to different subsets of the original cross-sectional sample of 3,049 women. The sample size that has been provided in Table 3 was the sum of the number of women in the four menopausal status groups to whom the measure of depressive symptoms (CES-D, see Chapter 3) was administered at two time points. Oddly, in their discussion, the NHEFS researchers attributed their failure to replicate McKinlay et al.'s (1987) cross-sectional "finding that depression is associated with surgical menopause" to the size of the Massachusetts sample: "Only 78 women experienced a surgical menopause during the initial follow-up interval (9-27 months)" (Busch et al., 1994, p. 223). According to the NHEFS's tabular presentation of results,

however, there were only 50 women in their own “surgical menopause” group to whom the CES-D was administered at two time points.

### Eligibility Criteria

Eligibility criteria for inclusion in the longitudinal analyses varied between the studies reviewed here. Several studies required women be “premenopausal,” however defined, at entry (Bromberger & Matthews, 1996b; Busch et al., 1994; Hunter, 1990, 1992; Kaufert et al., 1992). For entry to the Massachusetts study women were required to have menstruated within three months; however, they might have been postmenopausal by the time the first measure of depressive symptoms (CES-D) was administered, for some nine and for some 18 months later (Avis et al., 1994).

Bromberger & Matthews (1996b) specifically excluded women at entry who were not “healthy,” that is, who were taking antidepressants and/or other prescription medications or whose depression scores already indicated major depression at entry. They noted that their sample demonstrated a higher SES than non-participants, with higher educational attainment. This finding would be consistent with their requirements for healthy status at entry (Haan, Kaplan, & Camacho, 1987).

Most importantly, several studies specifically excluded women on HRT and/or after surgical menopause or hysterectomy (Avis et al., 1994; Dennerstein et al., 1997; Hunter, 1990; Holte, 1992; Hällström & Samuelsson, 1985). Although not explicitly stated, but based on their table of sample characteristics, Avis et al. (1994) apparently excluded women on HRT at entry to the longitudinal study, but retained them in the

sample if later they had used HRT. Rates of HRT use were therefore very low in their sample, 5.1% among women who had not had hysterectomies.

Yet, if the purpose of the study was to investigate the effect of menopausal or menstrual status or its change on depressive symptoms, eliminating portions of the sample due to use of HRT, for example, meant that the researcher lost one of the common intervening variables possibly related to depression, if the depressive symptoms were the reason why the women received HRT and/or hysterectomy. It is well established that many women with depression somatize and present with gynecologic complaints (Kirmayer & Robbins, 1996), the latter often leading to HRT prescription and/or hysterectomy in midlife (Gath, Cooper, & Day, 1982). It is also common practice for midlife women with depressive symptoms to be prescribed HRT (Derby, Hume, Lamont, Barbour, McPhillips, Lasater, & Carleton, 1993; Hemminki, Kennedy, Baum, & McKinlay, 1988; Klaiber, Broverman, Vogel, & Kobayashi, 1979; Strickler, Borth, Cechutti, Cookson, Harper, Potvin et al., 1977). Therefore, excluding those women from the sample created a bias in determining the relationship between menopausal status and depression.

In fact, the reason these nine studies were conducted was because of the common professional and popular belief that menopausal transition causes depression. Based on this belief the logical intervention is to treat the depression by treating the menopausal transition, namely by prescribing HRT. However, this practice means that studies that excluded women on HRT might have, in essence, excluded the portion of their outcome variable that they were most interested in. Elimination of the group using HRT was

particularly problematic when this group comprised a large minority of the original cross-sectional sample (Holte, 1992). Caution is warranted therefore in comparing results between studies and especially in generalizing them to all midlife women.

### Sociocultural Characteristics Of Samples

Sociocultural differences between countries and even between regions within the U.S. might also have contributed to differences in outcomes. While the focus of the present research was not cross-cultural, many other investigators have studied such differences (Beyene, 1986; Davis, 1986; George, 1988; Kaiser, 1990; Kay, Voda, Olivas, Rios, & Imle, 1982; Lock, 1986; Obermeyer, 2000; Punyahotra, Dennerstein, & Lehert, 1997). The nine longitudinal studies reviewed here compared in their discussions women living in Sweden, Norway, Australia, England, Canada, and various regions and the whole of the U. S. Assumptions of cross-country similarities were not made explicit but implied by Hunter (1993), in discussing five of the longitudinal studies also reviewed here, that women in all five countries were readily comparable.

One other mention of possible differences, and then not so much in the women themselves but in health services, was provided by Busch et al. (1994) in comparing their results from their representative national U. S. sample to those of the cross-sectional sample from Massachusetts alone (McKinlay et al., 1987). In this case, Busch and colleagues attributed a higher score for mean depressive symptoms found by the Massachusetts study in the “surgical menopause” group to the distinctive characteristics of this group in the Northeast U. S. Busch and colleagues noted that hysterectomy rates were low in the Northeast; then they explained that low rates indicated that these women

were more likely, therefore, to have had poorer health status than women having a surgical menopause elsewhere in the U.S. Further, women whose health status was poorer were more likely to be depressed. Busch et al. therefore concluded that the higher level of depressive symptoms among the surgical menopause group of the Massachusetts study might have been due to health status, of which the surgery was merely an indicator. The possibility that there might have been socio-cultural differences in the Northeast women was missing from their discussion.

Other examples of sociological differences included employment, education, and marital status. In 1991, the Melbourne study's sample of Australian born women aged 45-55 years demonstrated a 36% rate not employed for pay outside the home, and fewer than half (48.2%) had more than a 10<sup>th</sup> grade education. The Manitoba study's longitudinal sample also demonstrated a 41% rate not employed for pay outside the home and 33% with less than a secondary education. These women might not be comparable to those of the Seattle study where at entry only 14% were not employed for pay, and where the mean number of years of education was 16. For another example, Dennerstein, Smith, and Morse (1994) identified being not "married or living with a partner" as significantly related to depressive symptoms in their cross-sectional analysis. Differences in rates of marriage or cohabiting might affect differences in rates of depression or levels of depressive symptoms in other studies. What effects sociological and/or cultural differences between countries might have had on the results is unknown at this time.

Differences between samples also were evident in the proportions that had high depressive symptoms. From a low in the Gothenberg study (Hällström & Samuelsson, 1985) of 6% to a high of 28% in the Seattle study (Woods & Mitchell, 1996) is a more than fourfold difference in prevalence of what is interpreted as major depression.

Prevalence rates affect calculations of significance (Fleiss, 1981, Chap. 6).

### Cohort Effects

Cohort effects refer to bias or confounding based on historical changes in factors influencing the outcome variable (Chiazze, Brayer, Macisco, Jr., Parker, & Duffy, 1968 p. 90). For example, women born earlier in this century were less likely to have had paid employment than women born more recently. If paid employment affected depressive symptoms, then women born at different times might be expected to have different rates of depressive symptoms due to differences in rates of paid employment in different age groups (cohorts) (Bernard, 1981, Chap. 6). In this review, cohort effects would have been determined by the year in which the data were collected combined with the age of the women in that year. The earliest data collection occurred in 1968 in the Gothenburg study with women as old as 54 years then (Hällström & Samuelsson, 1985) while the latest occurred in Melbourne in 1994 with women as young as 49 years (Dennerstein et al., 1997) and in Seattle in 1991 with women as young as 35 years (Woods & Mitchell, 1996).

Cohort effects were identified in the Gothenburg study, probably because of the sampling methods providing an 18-year spread in age. Women were sampled from 38, 46, 50, and 54 year olds for the “first wave” of data collection in 1968. Cross-sectionally,

the women in the 38-year-old and 46-year-old cohorts had identical prevalences of mental disorder regardless of menopausal status. However, women in the 50- and 54-year-old cohorts had a nonsignificant trend “towards rising prevalence of mental disorder” the later the menopausal status. This finding was explained by further analyses: Women who were older at the first wave were more likely to have had a mental disorder when they were 30-39 years old than the 38-year-old women at the age of 30-39 years. The incidence (onset of new disease) of mental disorder was the same across ages and menopausal status. Therefore, when age at first wave was adjusted for, there were no differences in disability from mental disorder. Because differences in mental disorder and functional disability were no longer significant after adjusting for age, the Gothenburg researchers concluded there was a cohort effect on mental disorders.

The cohort effects were not straightforward, however. Although age was a confounder for prevalence of mental disorder at ages 30-39 years, the diagnoses of mental disorder and functional disability for the period of ages 30-39 years were made based on retrospective recollection of symptoms. Thus, recall times were longer for the women who were older at the first wave than for the younger women, although recall bias was not identified as such by Hällström and Samuelsson (1985). Complicating the matter further, participation in the study increased with increasing age cohort. The Seattle study also entered women with a 20-year age spread. What further significance cohort effects might have had between the nine studies reviewed here was unknown at the time of this review.

### Measurement Tools

All of the researchers used self-reported symptom measures of depression except Hällström and Samuelsson in Gothenburg, Sweden. Some of the self-reports were done by mail questionnaire, some by telephone or personal interview. The most frequently used tool (see Table 3) was the Center for Epidemiologic Studies-Depression (CES-D) Scale (Radloff, 1977). Other tools included standardized ones such as the Beck Depression Inventory (BDI), used in Pittsburgh (Bromberger & Matthews, 1996b), an adaptation of the General Health Questionnaire (GHQ) used in Norway (Holte, 1992), and an adaptation of the Affectometer-2 used in Melbourne (Dennerstein et al., 1997). Through principal components analysis Hunter developed a measure of depression from a set of 36 symptoms from the Women's Health Questionnaire (Hunter, 1992).

Self-report measures of depressive symptoms were developed for ease of administration in research studies and were "designed to measure current level of depressive symptomatology, with emphasis on the affective component, depressed mood" (Radloff, 1977, p. 385). Hällström and Samuelsson in Gothenburg, Sweden (1985) were the only researchers to conduct psychiatric interviews to establish actual diagnoses of mental health problems. They used a semi-structured interview that lasted from one-half to two hours (Samuelsson, 1982). Unfortunately, a potential bias was introduced by their procedure of one researcher (Hällström) conducting all of the interviews for the first wave and the other (Samuelsson) conducting all of the interviews for the second wave. Their inter-rater reliabilities were generally high, but in the one

area of psychiatrist-observed behavior, inter-rate reliabilities were so low as to be unusable (Samuelsson, 1982).

Unfortunately also, these authors generally reported on the entire category of mental health problems without separating out depression. However, these authors did provide a breakdown of specific diagnoses within the larger category of mental health problems cross-sectionally for one time point, first wave. They pointed out that the single diagnosis of major depression was more than 50% of all their mental health diagnoses. A diagnosis of “neurasthenic state,” not further defined but possibly similar to dysthymia, was their third most prevalent one, applying to 1.8% of their total sample. However, this diagnosis is not present in the DSM-IV (1994). While some of the procedural elements (reporting all diagnoses as a single group, changing interviewer) might be problematic, these researchers were the only ones who measured the diagnosis of major depression according to criteria of the DSM-III, including the level of functional disability. For ease of writing, the phrase “depressive symptoms” was used to refer to all nine studies, with the understanding that the Gothenburg women diagnosed with “mental health problems” in the majority had major depression and therefore were, in the majority, experiencing depressive symptoms.

#### Menopausal And Menstrual Status

While the variable of menopausal and menstrual status was only one of several that were included in the nine studies, it was isolated for review here because of its central importance for longitudinal studies of menopause and depression. Conceptual issues in defining menopausal status were reviewed in 1987 by Kaufert, Gilbert and Tate

. As they pointed out, commonly used categories of menopausal status were derived from only a limited amount of empirical information, particularly scant hormonal information but also little cycle and flow data as well. For example, in their own longitudinal analysis, Kaufert et al. (1987) started with, but provided no particular basis for, a menopausal status of amenorrhea for three months and in fact ended up combining this category with irregular cycles. In addition, according to their own data (and replicated in Holte's study) amenorrhea for six months, initially a separate category, was actually the beginning of postmenopause in over 90% of women, who continued on to 12 months of amenorrhea. Their purpose in combining irregular menstruation with amenorrhea for the previous three months was "to simplify presentation" (Kaufert et al., 1987, p. 220). Their purpose in separating the category of "amenorrhea for the last 3 months" from "amenorrhea for the last 6 months" was not provided. The limited empirical basis for definitions of menopausal status creates the potential for uncertainty in classification, as evidenced by the Manitoba researchers, and even misclassification.

#### World Health Organization Definitions

The World Health Organization (WHO) has proposed on at least two occasions standardizing terminology in menopausal research (1981, 1996), with less than universal acceptance. A quick review of their definitions as most recently re-proposed (1996) is provided here.

Menopause was defined as the final menstrual period (FMP) of a woman's life. The report also mentioned that it requires 12 months of amenorrhea to recognize a menstrual period as the final one of a woman's life. The WHO Report included the

comment that natural menopause is understood to result “from the loss of ovarian follicular activity” and that there should be “no other obvious pathological or physiological cause” for the amenorrhea (1996, p. 12-13). Presumably other causes could be identified in some women. In their early article, Kaufert et al. (1987) specifically discussed their failure to ask about alternative causes of irregularity or amenorrhea other than menopausal transition.

The WHO Report did not clarify that women sometimes experience vaginal bleeding or spotting at the end of menstrual life (with no other pathological or physiological cause) that the women, themselves, do not recognize as a FMP because the flow is lighter, and/or a different color, and/or of shorter duration than they have been used to. This circumstance did in fact caused some confusion in the SMWHS about the exact date of menopause. For example, the WHO Report did not caution researchers against using questions such as, “Have you had a period in the last xx months?” as opposed to, “Have you had any episodes of vaginal bleeding and/or spotting in the last xx months, whether they were your period or not?”

According to the WHO Report, menopausal transition “should be reserved for that period of time before the FMP when variability in the menstrual cycle is usually increased” (p. 13). *Perimenopause* was recognized to have quite varying meanings in the research literature, but recommendation was made for perimenopause to include all of menopausal transition plus the first 12 months immediately postmenopause. Perimenopause therefore included the time required to define menopause as such.

*Postmenopause* was defined as starting at the date of the FMP. The WHO Report was explicitly clear that the first 12 months after the FMP were included in postmenopause. There is, therefore, an overlap with perimenopause. For research purposes, therefore, the WHO definitions for perimenopause and postmenopause are difficult to use in the same study because they are not mutually exclusive. Furthermore, particularly in prospective data collection, the first 12 months after the FMP cannot be identified as such at the time. The FMP, and thus the start of postmenopause, can be identified only in retrospect. Therefore, some researchers define postmenopause as starting after the perimenopause, that is, starting after the 12 defining months of amenorrhea after the FMP, although this definition is not in agreement with the recommendations of the WHO Report. In order to be consistent with WHO recommendations, researchers could adopt a term “post-perimenopause” to indicate the time after the 12 defining months of amenorrhea post FMP.

*Premenopause* was recognized as having variable meanings, although the recommendation to define it as “the whole of the reproductive period prior to the menopause” (p. 13) made it so broad as to be unhelpful. The WHO definition of premenopause, therefore, created an overlap in meaning between premenopause and perimenopause, as well as menopausal transition. The WHO Report did acknowledge that many researchers limit the meaning to the “one or two years immediately before the menopause” (p. 13) although they did not cite any such authors.

### Evidence-Based Definitions

Before the SMWHS, few researchers (Kaufert et al., 1987; Treloar, 1974; Vollman, 1956, 1977) used prospectively collected menstrual cycle information to characterize the transition to menopause. They focused on menstrual cycle length and variability, although the first two also reported on flow duration. It is Treloar's and Vollman's description of increasing variability of cycle length that is the basis of the WHO definition of menopausal transition (WHO, 1996, p. 13), which Treloar labeled "perimenopausal."

Kaufert et al. (1987) advanced Treloar's and Vollman's observations by noting that "viewed from a longitudinal perspective, the assumptions underlying the accepted definition of menopausal status do require some revision" (p. 225). The Canadian researchers emphasized that not all women followed a strict developmental model of linear progression through phases, even for the three years of their study where women were interviewed by telephone every six months. "Reverse changes, such as the return of menstruation or a move from irregularity to regularity" occurred (Kaufert et al., 1987, p. 222). They also cautioned against "thinking of menopausal status as 3 categories which are internally homogenous and have sharply demarcated boundaries" (p. 225), the three categories being pre-, peri- and postmenopausal. They noted that cross-sectionally a woman who might be classified as perimenopausal based on irregularity within the previous designated number of months might return to regularity, or that a woman who reports amenorrhea for the previous three or six months might menstruate again or might

have achieved menopause with that last menstruation. The amount of time in each phase varied, as well.

Mitchell, Woods, & Mariella (2000) were the first researchers to identify evidence-based stages *within* the menopausal transition. The characterization of these stages then helped to clarify the beginning of the menopausal transition phase, that is, the time before the onset of the menopausal transition in midlife women. Four characteristics of menstruation were identified from intensive review of calendar data, from responses by participating women to yearly open-ended questions about changes within the previous 12 months, plus from a telephone interview. The four characteristics that changed during the menopausal transition were: 1) menstrual flow amount 2) flow duration 3) cycle length and 4) cycle regularity. One note on terminology: For Mitchell et al. (2000) the term *phase* of menopausal transition referred to the larger categories of pretransition, menopausal transition, and postmenopause. *Stage* of menopausal transition referred to the subcategories within the phase of menopausal transition. However, in this review all divisions have been referred to as menopausal transition stage or MTS.

Pretransition was defined as the midlife years (from age 35 years) before any flow or cycle-length changes occurred. These cycles and menses are unchanged from those during the woman's 20s and early 30s in the four menstrual characteristics, duration of flow, amount of flow, cycle length, and cycle regularity. Menopausal transition was defined as starting at the onset of those changes that occur after age 35 years and that persist from one year to the next. Transition continues to the FMP, that is, to menopause. Postmenopause was defined as starting from the FMP and is recognized after 12 months

of amenorrhea. Postmenopause continues until the end of all ovarian-hormone-related changes, which is an indefinite time period.

Menopausal transition was then further divided into early, middle, and late stages. Early transition, a stage heretofore unrecognized in the research literature, was characterized by changes in at least one of three menstrual characteristics: flow amount, flow duration, and (still regular) cycle length. In the Seattle study, 37% of women had experienced at least one of these changes, that is had entered early transition, before age 40 years. An additional 41% (for a total of 78%) had experienced at least one change before the age of 45 years. Thus, limiting sampling to women 45 years or older as was done in the Massachusetts, Manitoba, Melbourne, South-East England, and Oslo studies, missed the effects, if any, of early transition for the majority of women.

Middle transition in the SMWHS was characterized by new irregularity that persisted across years, defined as at least one episode of >6 days difference in consecutive cycle lengths for women who did not previously experience a difference this large. Irregularity was found to precede skipping on the average by two or more years. The specific designation of >6 days difference was based on prospectively collected menstrual cycle data from regularly menstruating women (Chiazze et al., 1968; Mitchell et al., 2000; Treloar, Boynton, Behn, & Brown, 1967).

Skipping became the distinguishing characteristic of late transition. Skipping was defined as no bleeding or spotting for a time greater than or equal to twice the modal cycle length for that calendar year. Other researchers' definitions of perimenopause that specify a certain number of months of amenorrhea might be seen as corresponding to late

transition, although specifying a particular time in “months” meant the definition was somewhat imprecise. While middle transition might have coincided with Treloar’s definition and therefore the WHO Report definition, the distinguishing of late transition as a later phase had also been missing in previous literature.

The time frame used to establish the presence of the characteristics that define each stage was a calendar year. The use of a year as a time frame was arbitrary. Mitchell et al. (2000) observed, similar to Kaufert et al. (1987) that even though there might be a year that showed at least one cycle of either irregularity or skipping, regular cycles might occur at other times during that year, or even in subsequent years. Again similar to the sample studied by Kaufert and colleagues, more than half of the women in the SMWHS did not exhibit a linear forward progression but instead reverted to a previous stage after one or more years at a later stage.

Later researchers (McKinlay et al., 1992) used Treloar’s observations as a partial basis for their definitions of premenopausal, perimenopausal, and postmenopausal. At the same time that these Massachusetts researchers were proposing their study, the Manitoba project was developed independently along similar lines (Kaufert, 1984; Kaufert et al., 1992). The two projects informally agreed to use similar definitions. Then Dennerstein and colleagues in their longitudinal study of depressive symptoms deliberately followed the lead of the Massachusetts and Manitoba studies in order to allow for comparisons across studies (Burger, Dudley, Hopper, Shelley, Green, Smith et al., 1995; Dennerstein et al., 1997).

Unfortunately, the WHO Report (1996) cited McKinlay et al. (1992), asserting that the Massachusetts researchers had “clarified the concept of menopausal transition” (p. 14). These researchers arbitrarily limited their sample to women 45 years or older and pre-defined menopausal transition’s “lower limit [as] the onset of menstrual cycle irregularity.” Irregularity was not defined. Then, according to the Massachusetts researchers, average duration of the menopausal transition was 3.8 years. However, when menopausal transition was pre-defined as irregularity in women 45 years or older, other characteristics of transition such as those identified by Mitchell et al. (2000), and those changes in women younger than 45 years were omitted a priori. It was possible, however, that what McKinlay et al. (1992) observed was the duration of an approximation of late transition that started on average at 47.5 years of age and lasted an average of 3.8 years.

#### Comparisons Of Definitions And Implications

In the nine longitudinal studies, the phrase *menopausal status* was the most commonly used one to label the variable of interest here. The more recent phrase, *menopausal transition*, was also used occasionally, and was more likely to emphasize the changes or transition that women experience during the developmental change from a fertility-oriented menstrual cycling phase to a post-fertility post-menopausal phase of their lives (Fogel & Woods, 1995). The phrase *climacteric* has also been used to emphasize transition (Magursky, Mesko et al., 1975; Pritchard & MacDonald, 1976, p. 74; Utian, 1994). Magursky et al., (1975) and Utian (1991, 1994) also understood the term climacteric or climacteric syndrome to require *symptoms* of menopausal transition.

Table 2 has provided a summary of definitions of menopausal and menstrual status as used in the nine studies reviewed here. As can be seen from inspection of Table 2, many of the labels and some of the definitions used for menopausal status categories followed the recommendations of the 1996 WHO Report. Some of the menopausal status categories were also similar to those Mitchell et al. (2000) identified. Many were arbitrary and conflicted with each other. The next section entails a comparison of label use and definition between the nine longitudinal studies of depressed mood in midlife women. The stages of transition from Mitchell et al. (2000) were used as the standard for definition and labeling.

Woods & Mitchell's (1996) analysis of depressive symptoms measured at two time points one year apart used a longitudinal time frame of "the last six months" for menstrual changes. This article was published before the stages of menopausal transition were specifically defined. Woods and Mitchell assessed changes in regularity and in the interval between menses because "changes in regularity and cycle length have been associated with menopausal transition" (1996, p. 115). It must be assumed that a variable "menopausal transition," defined by changes in regularity and/or cycle length, was used. The dearth of information in the published article, however, made it difficult to know if menopausal transition was compared to pretransition and/or postmenopause.

Holte (1992) used definitions that made comparison with other studies less than satisfactory. He uncommonly defined menopause as six months of amenorrhea instead of 12. Then, the few women having recurring menses after six months of amenorrhea were excluded from the analysis. This biased that sample by excluding the very women in late

transition who might be experiencing changes in their mood related to prolonged menopausal transition, the group in which the Massachusetts study found increased depression scores (see Results section later). Holte justified his definition of menopause by claiming that 12 months was “possibly insensitive” but provided no further elaboration of his thinking. The reader might speculate that Holte assumed some special effect produced at the moment of the FMP, and that waiting 12 months to identify the FMP might have dissipated the effect. Because of the short, one-year time frame for menstrual changes in Woods & Mitchell (1996) and the unusual use of the menopausal status variable in Holte (1992), critical review of the menopausal status variable was limited to the seven remaining studies; review has been organized by MTS.

Preclimacteric: Hällström & Samuelsson (1985) were the only investigators who labeled a MTS as preclimacteric. In Gothenburg this category was determined by age only, women who were 38 years old at entry to the study. Woods & Mitchell’s sample also included women less than 40 years old at entry to the study, their lowest age at entry being 35 years. The stage of pretransition might correspond to the stage of preclimacteric in the Gothenburg study, although the latter would apply at the same time to women who had entered early transition before age 40 years.

Premenopausal: The premenopausal category was similar although not identical between studies, and usually included the criteria of regularity and/or lack of change. These definitions did not usually conform to the WHO Report recommendations defining premenopause as the time from menarche to the FMP. Except for Avis et al. (1994) and Dennerstein et al. (1997), it was assumed that the aspect of change considered was that of

(ir)regularity alone, not flow amount, flow duration, or change in regular cycle length. Avis et al. and Dennerstein et al. did specify changes in flow as well as cycle length.

The premenopausal category as used by researchers might more accurately have been labeled pretransition, or at least preceding what was generally understood to be the menopausal transition. The concept of change might be an important co-characteristic with regularity, however, since not all women less than 40 or even less than 35 years old have regular menses to begin with (Treloar et al., 1967). For these women, unchanged irregularity at age 47 would be classified as pretransition by Mitchell et al. (2000). Dennerstein et al. (1997) and Woods and Mitchell (1996) were the only authors who specifically mentioned the characteristic of *no change* as part of the definition of this category.

Unfortunately, different time frames were used, varying from any menses in the last three months (Bromberger & Matthews, 1996b) to regular menses for the last 12 months (Hunter, 1990). Because women do not necessarily progress through the menopausal transition linearly, Bromberger & Matthews' definition of *premenopause* (any menses in last three months, see Table 2) or Avis et al.'s (1994) definition (regular menstrual periods the three months immediately before the interview, see Table 2) might well have included women in late transition who by chance had had at least one or three periods, respectively, in the months before the interview. These women, having already experienced the characteristics of late transition (skipping of menses) in months previous to the three months arbitrarily specified just before the interview, would be classified in premenopausal categories by Bromberger and Matthews and Avis and colleagues. For

Bromberger and Matthews, the one period a woman might have had within the last three months might have been her final period, so premenopausal in the University of Pittsburgh study also might have included women within three months after the FMP itself. In this case, the definition of premenopause used by Bromberger and Matthews could be interpreted as consistent with WHO's recommendations, that is, premenopause continues up until the FMP. However, the rationale for using a time frame of three months instead of 12 was left to the imagination of the reader.

Hunter's definition of *perimenopause* as a cycle length greater than six weeks would classify as *premenopausal* a woman whose usual cycle length was regular around 24 days but whose year included a cycle 35 days (5 weeks) long, a woman who would be in middle transition according to Mitchell et al. (2000). Hällström and Samuelsson (1985) defined regular menstruation as no amenorrhea for two or more months. According to Mitchell et al. (2000) skipping was a marker of late transition and was preceded by more than two years of less extreme irregularity. The reader could assume, therefore, that the Gothenburg study's definition of irregularity probably meant amenorrhea  $\geq 2$  months, that is, skipping or late transition rather than irregularity as defined by Mitchell et al. (2000), the defining characteristic middle transition.

The Massachusetts study (Avis et al., 1994) used different time frames to determine premenopause (3 months before the interview) and perimenopause (12 months before the interview). Again, given the lack of forward progression for many women found in both the studies by Mitchell et al. (2000) and Kaufert et al. (1987), it would be possible for a woman to have had regular menses during the previous three months and

still have had irregularity and/or changes in flow during the previous 12 months as well. This study, therefore, used definitions of premenopause and perimenopause that were not mutually exclusive. In summary, as can be seen, the category of premenopause was the most variable in definition and therefore created the most difficulty with comparisons across studies.

Perimenopausal: Common defining characteristics of perimenopause included the idea of irregularity and/or amenorrhea for a specified number of months, and occasionally changes in menstrual flow amount and duration (Avis et al, 1994; Hällström & Samuelsson, 1985; Hunter, 1992; Kaufert et al., 1992). However, irregularity was not defined by most of those authors reviewed here, and might have been understood by both researchers and/or participants to mean cycle length irregularity, cycle skipping or amenorrhea, flow irregularity and/or duration irregularity. The rationale for any specified number of months of amenorrhea was also universally lacking.

The NHEFS (1996) used such a general definition of perimenopause, irregularity due to the change of life, that it included all three stages of menopausal transition; however, since flow changes were not specified as irregularity, it was possible that Mitchell et al.'s (2000) early transition was categorized as premenopausal. Similarly, the Manitoba study (1992), after specifying premenopause as regular and postmenopause as 12 months of amenorrhea, used the term perimenopause to apply to women who by default were not pre- or postmenopause. Thus perimenopause probably included middle and late transition plus the first 12 months after the FMP. In the Manitoba study perimenopause was, however, mutually exclusive with pre- and postmenopause.

The Melbourne study was the only one to create two stages and adopt the labels early and late perimenopause. Late perimenopause was specified as 3-11 months of amenorrhea. This classification therefore included women during some of the immediate 12 months postmenopause plus some women in late transition. Early perimenopause was defined as some change in frequency and therefore included women in early, middle, and some in late transition. Women in late transition, therefore, could have been categorized into either early or late perimenopause. However, in an earlier cross-sectional article (Burger et al., 1995) the Melbourne researchers tried to sequentialize menstrual flow changes as first and irregularity of cycle length as second, which could correspond to early and middle transition. Such a classification could have identified early transition as a category. Nevertheless, they did not maintain these definitions for the longitudinal portion of their study.

While Hunter ended up combining peri- and postmenopause into one group for analysis, she started by defining perimenopause as irregular menses that she specified as a cycle greater than six weeks long. No basis or rationale for specifying six weeks was given. Women who did not have regular menses in the last 12 months and who also did not yet have six-week long cycle lengths would have been unclassifiable. Presumably because her definition of postmenopause did not start until a woman had 12 months of amenorrhea, her perimenopause category included women during the first year postmenopause. By combining peri- and postmenopause, Hunter presumably theorized that changes in depressive symptoms would be due to the change from pretransition to

menopausal transition, that is, due to the first menstrual changes, rather than around menopause.

Bromberger & Matthews (1996b), Dennerstein et al. (1997), and Holte (1992) might be interpreted as trying to prospectively capture the immediate months around the FMP in their stage labeled perimenopause. Yet there is no explicit rationale to expect psychological or mood changes during this narrow time span (Greene, 1992). There is also consistent empiric information that the menstrual changes of the menopausal transition take place over a much longer time frame, on the average years, not months (Kaufert et al., 1987; Treloar et al., 1967; Vollman, 1956). Definitions of perimenopause and designs of studies that eliminated the phase menopausal transition (Bromberger & Matthews, 1996b; Holte, 1992; Hunter, 1990, 1992) limited the ability to identify when depressive symptoms might change.

In summary, perimenopause contained some of the greatest variability in definition, even so far as to be a category by exclusion for Kaufert et al. (1992) where women were categorized as perimenopausal if they did not meet the criteria for pre- or postmenopause. Furthermore, Bromberger & Matthews (1996b) combined perimenopause with postmenopause into one category for longitudinal analysis, based on their failure to find differences between these groups in their cross-sectional analyses, published earlier (Matthews et al., 1990). Hunter (1990, 1992) also combined peri- and postmenopause. The potential for misclassification between the category of perimenopause and premenopause, especially by Avis et al. (1994), made interpretation and comparison of results imprecise. Since menopausal and menstrual status was critical

to understanding how menopause and the menopausal transition might be related to depressive symptoms, and especially since Avis et al. (1994), Dennerstein et al. (1997), and Hunter (1990, 1992) found that perimenopause was where mean or prevalence of depressive symptoms were higher, the problems in definition were important, potentially biasing results (see Results section later).

Postmenopausal: The differences between studies in defining postmenopause were related to the classification of the first 12 months after the FMP. As already mentioned, the WHO (1996) recommendation allows overlap of this year between both perimenopause and postmenopause. As already reviewed, Bromberger & Matthews (1996b) and Hunter (1990, 1992) obviated the problem of overlap although lost the ability to distinguish any effects associated with the FMP itself.

It must be assumed that other researchers defined the two terms of perimenopause and postmenopause as mutually exclusive. The Massachusetts study (Avis et al., 1994) defined natural menopause as no menses in the last 12 months, apparently starting postmenopause one year after the FMP. Likewise, the NHEFS (Busch et al., 1994), the Manitoba Project (Kaufert et al., 1992), Melbourne (Dennerstein et al., 1997), and South-East England studies (Hunter, 1990) included the first 12 months after menopause in perimenopause only. The Norwegian (1992), and Gothenburg (1985) studies matched the WHO Report recommendation and started postmenopause at the FMP. Holte's unusual determination of menopause has been previously discussed.

Postmenopause was divided into early and late postmenopause by a few authors, although little evidence-based rationale was provided for the various postmenopausal

categories. Changes during the first several years postmenopause might be intriguing to study. However, comparison and interpretation of studies would benefit from standardized definitions so that any differences found (or their lack) between the first, second, third, fourth, and/or fifth postmenopausal year could be determined as reliable. It would also be helpful to establish what changes would have been expected to occur between the first and second years of amenorrhea (Dennerstein et al, 1997), and/or between the second and third (Hällström & Samuelsson, 1985). While there might be differences within the first years postmenopause, Kaufert & Syrotuik (1981) found that vasomotor symptoms persisted and even increased slightly in the group five or more years postmenopause compared to those less than five years postmenopause.

Surgical Menopause: Additional menopausal status categories included surgical menopause, or hysterectomy. While the removal of the uterus causes menses to stop, the categories of menopausal status were usually understood to reflect underlying changes in ovarian and hypothalamic-pituitary hormone cycling. In that context it would have been valuable to develop categories based on the presence of at least one ovary, rather than the absence of the uterus. However, after hysterectomy any hormonal cycling or lack thereof would not be identifiable from external evidence such as menses. In the few studies where they were not excluded, mixing of women in this category who had bilateral oophorectomy with those who did not (Busch et al., 1994; Kaufert et al., 1992) and/or were on HRT (Bromberger & Matthews, 1996b) made this category too mixed to compare results between studies.

**Hormone Replacement Therapy (HRT):** All researchers identified women using hormone or estrogen replacement therapy (HRT or ERT), either excluding them or separating them into their own category. HRT is used for women with an intact uterus and ERT is used after hysterectomy. HRT includes a progestin component in order to decrease the risk of endometrial cancer that ERT alone entails. However, the progestin component in and of itself has been implicated in various symptoms, in particular depressed mood, although evidence of progestin-related symptoms is inconsistent (Westhoff, Truman, Kalmuss, Cushman, Davidson, Ruline et al., 1998; Westhoff, Truman, Kalmuss, Cushman, Ruline, Heartwell, et al. 1998), but still a potential confounder.

Theoretically, use of HRT is an independent category unrelated to MTS. For example, a woman might be in middle transition on HRT, in late transition on HRT, or postmenopausal on HRT. However, once a woman starts HRT, determination of her MTS is impossible unless she started HRT more than 12 months after her FMP. In practice, therefore, HRT use becomes mutually exclusive with the other MTS and has been included as a category of MTS. At this time it is unknown whether a researcher should consider HRT use as a single (homogeneous) category or if further subdivisions or another approach would be warranted. The SMWHS has attempted to maintain some distinction by maintaining two HRT categories, one of definitively postmenopausal women based on  $\geq 12$  months of amenorrhea before starting therapy and one where the menopausal status is unknown. With the recent practice of prescribing HRT for women before menopause (Derby et al., 1993; Hemminki et al, 1988; Rueda Martinez de Santos,

1997) it remains to be seen how women using HRT should best be categorized in research.

In spite of its central importance to all of the studies reviewed here, in all nine studies menstrual cycle and flow information were obtained by retrospective self-report only. While women may be quite aware of changes in their menstrual *flow*, and remember these changes in detail for some time, self reports about menstrual *cycles* and changes in cycle length are less reliable (Treloar et al., 1967). In fact, during the course of the SMWHS, it has been observed that prospective recording of menstrual cycles on calendars bears only a modest relationship to one-year retrospective self reports on an annual questionnaire (Ellen S. Mitchell, paper presented at the Society for Menstrual Cycle Research, 1997, in press). Use of prospectively recorded information about bleeding and flow on calendars might be required for more accurate categorization of women regarding menopausal and menstrual status.

Measurement of MTS relates to research design. Most of the researchers reviewed here seemed to assume that a particular status, that is, being perimenopausal or being postmenopausal (however defined) might relate significantly to depressive symptoms. However, as Kaufert et al. pointed out (1987) “researchers cannot discriminate between women close to a change in their menopausal status and those whose status will remain relatively stable for a year or more” (p. 218). The reason researchers cannot discriminate, as discussed earlier, is that the changes of menopausal transition do not necessarily occur in a forward linear progression. In addition, as Mitchell et al. (2000) have described, each stage of the transition lasts on average two or

three years; however, an individual woman might experience a shorter or longer stage than the average. Thus for the researcher, the first months of perimenopause will be classified similarly to the last months, the latter perhaps being three years later, at the FMP.

Avis et al. (1994), Kaufert et al. (1992) and Bush et al. (1994) used change in status as the factor that might be related to depressive symptoms. For the latter group, the three categories were women who (after being selected as premenopausal at time one) had stayed premenopausal, had become perimenopausal or had become postmenopausal at time two. The former groups used similar paired-status categories. Avis et al. (1994) found another aspect of status, that is, the duration of time within perimenopause, as significant. It might be important for future researchers to be able to consider all three aspects of menopausal status: being in a stage, changing from or to a stage, and duration of time in a particular stage of the menopausal transition.

#### Other Independent Variables

As was explained by several researchers, theories and research about depression indicate that other areas of women's lives must also be measured in order to help interpret measures of depressive symptoms (Kaufert, 1984; Kaufert, Lock, McKinlay, Beyenne, Cooper, David et al., 1986; Woods & Mitchell, 1996; Woods & Mitchell, 1997). Holte (1992) pointed out that failure to control for confounders such as age, life events, hormonal status (i.e., use of HRT), and previous premenopausal complaints has made results of many studies uninterpretable. While each author used conceptual labels to discuss other independent variables included in their analyses besides depressive

symptoms and menopausal status, the conceptual labels varied greatly from study to study. In order to facilitate comparisons across studies here, the concepts have been disassembled into individual measures or variables. The two exceptions to this approach were Seattle (Woods & Mitchell, 1996) and South-East England (Hunter, 1990). For Woods & Mitchell the large number of other independent variables warranted a factor analysis to reduce the richness into a usable number. Hunter also used factor analysis and subscales from a large list of items. For the Seattle and South-East England studies, therefore, the factor or subscale names have been used.

Other independent variables ranged from so many that data reduction through factor analysis was warranted (Woods & Mitchell, 1996), to simply two by Busch et al. (1994), timing of menopause (one of three age categories at menopause—early, on-time, and late), and years since menopause. In the middle ground were the remaining seven studies. Because the NHEFS (1994) did not use other variables in its analysis, this discussion of other independent variables, except for age, has not mentioned that study further.

### Life Events and Stressors

Life events, stressful life events, stressors or stress have been some of the main variables of interest in causal theories of depression (Brown, 1998; Dohrenwend, 1998). Four studies included some measure of what may be labeled stress or life event stressors. The definition and measurement of this variable ranged from a single yes/no question about current stressfulness of life (Hunter, 1990) to complex standardized scales (Bromberger & Matthews, 1996b; Woods & Mitchell, 1996) to complex self-developed

scales (Hällström & Samuelsson, 1985). Some of the intermediate measures of stressors, as well as the complex scales, often categorized stressors into conceptual areas such as family, work, deaths, relationships. While some studies found one particular category significant in determining depressive symptoms, there was no consistency between studies in which categories of stressors were related to higher depressive symptoms.

Three of the researchers distinguished between current life stressors and recent history (within various specified time frames) of stressors and treated these as separate variables (Bromberger & Matthews, 1996b; Hällström & Samuelsson, 1985; Kaufert et al., 1992). None of these researchers provided a rationale for distinguishing current from recent history of stressors, nor did they provide a rationale for any particular time frame to define recent stressors in terms of effects expected on depressive symptoms. Nuckolls et al. (1972) proposed several decades ago that effects of stress might be buffered or modified by social support. Woods & Mitchell (1996) were the only researchers to measure resources and support, as buffers or modifiers of stress. All measurements of stress or stressors were by self-report. As discussed in the earlier section on causal models of depression, concurrent measurements of depressive symptoms, general symptoms, social support, stress, and life events by self report are likely to be correlated, thus confounding analysis.

For some researchers, menopausal transition itself was conceptualized as a life stressor. Bromberger & Matthews (1996b) considered it one of the three major subsets of life stress, along with life events within six months and current chronic stressors. Woods & Mitchell (1996, p. 115) used a measure of stress, “Norbeck’s adaptation of the Sarason

Life Events Survey,” where one of over 100 items of possible stressors was “menopause started.” Hällström & Samuelsson (1985) also included one item “menstruation ceases” in their weighed 52-item life events measure, although its weight was the lowest of all 52.

### Age

Age or age within MTS was another commonly measured covariate, although there is an important difference. Age is related to menopausal status; older midlife women are more likely to be at a later stage of menopausal transition than younger midlife women (Kaufert et al., 1987). When two variables are collinear, assumptions of regression have been violated, namely that the two variables be independent, not correlated (Kleinbaum et al., 1988, p. 206-214). Therefore, just adding age to a model predicting depressive symptoms that also uses menopausal status would violate statistical assumptions. However, adjusting MTS for age would allow both variables to be in the model appropriately. The question is whether or not there is variability of age within the stage that might be meaningful. For example, younger perimenopausal women (however defined) compared to older perimenopausal women might have experienced more depressive symptoms. Also, perhaps the lack of standardization of the perimenopause category might be compensated for by adjusting for age within the category.

Avis et al. (1994), Bromberger and Matthews (1996b), and Woods and Mitchell (1996) did not include consideration of age in their longitudinal analyses. How Kaufert et al. (1992) included age in their multiple logistic regression was not reported. The other researchers either repeated analyses adding age to an analysis of variance (Busch, et al, 1994; Holte, 1992) or used statistical methods such as analysis of covariance to assess

age effects (Hunter, 1990) or adjusted for age within MTS (Busch et al., 1994; Dennerstein et al., 1997; Hällström & Samuelsson, 1985; Holte, 1992; Hunter, 1990). None of the researchers who considered age found any significant relationship of age to depressive symptoms.

### Vasomotor Symptoms

Vasomotor symptoms such as hot flashes, hot flushes, and day or night sweats were also included as an independent variable by Woods & Mitchell (1996), Avis et al. (1994), and Dennerstein et al. (1997). Avis et al. (1994) discussed, however, that “symptoms and depression were both measured by self-report and may well be manifestations of the same underlying process” (p. 218). For the other researchers, there was presumably the assumption that physical symptoms might have caused depression. Avis et al. have detailed various mechanisms, such as a direct effect of vasomotor symptoms on depressive symptoms and an indirect effect. In the latter mechanism, vasomotor symptoms cause sleep disturbance, leading to sleep deprivation and thereby depressed mood (Baker et al., 1997). Levels of vasomotor symptoms are associated with MTS (Morse et al., 1994). Similar to age, therefore, vasomotor symptoms and MTS might be collinear if both were entered into a regression model. In spite of having the largest sample of women, Avis et al. (1994) also were unable to include in their model more than a single variable that combined hot flashes, night sweats and “menstrual problems,” because not enough women in various menopausal status groups experienced them. While menstrual problems were not defined by Avis and colleagues, they cite Magursky et al. (1975) who explained that these problems were “disorders such as oligo-,

hypo-, hyper-, or polymenorrhoea” (1975, p. 18). These “problems” can be understood as the defining criteria for early and middle transition (Mitchell et al., 2000).

### Marital Status

Marital status was included as an independent variable by four researchers (Avis et al., 1994; Hällström & Samuelsson, 1985; Hunter, 1990; Woods & Mitchell, 1996). For Woods & Mitchell marital status was only one item loading on the factor labeled Family Resources. Other items that loaded on this same factor were income, number of children, and the woman’s current employment. Because Hällström & Samuelsson also included some of these variables but did not present information on correlations between them, it cannot be known whether or not each one of these concepts always belonged in the model explaining depressive symptoms or whether they were, in fact, elements of a single factor such as Woods & Mitchell (1996) found. The definition of married might have varied between studies as well. Woods & Mitchell (1996) combined married and partnered, as did Hällström & Samuelsson (1985). Details of definition of marital status were not available for the other studies.

In spite of the difficulties in comparing use of the other independent variables of stressful life events, age, vasomotor symptoms, and marital status in various investigations, those were four of the independent variables more consistently used. Continued perusal of other independent variables made comparisons between studies even more difficult.

### Socioeconomic Status (SES) and/or Other Demographics

Various socioeconomic and/or demographic factors other than marital status have been identified in the literature as associated with depression, such as lack of employment outside of homemaking (Kandel et al., 1985; Miller, Wilbur, Montgomery & Chandler, 1998; Nathanson, 1980; Wilbur & Dan, 1989). Avis et al. (1994), Bromberger & Matthews (1996b), Hällström & Samuelsson (1985), Hunter (1990), and Woods & Mitchell (1996) included some measurement of education, income, occupation, and/or parity, as well as marital status, as covariates and predictors of depressive symptoms. Dennerstein et al. (1997) described their longitudinal sample in terms of parity, paid employment, and living with a partner but had not found these variables to be significant in an ANOVA of a factor labeled “dysphoria” from cross-sectional data (Dennerstein et al., 1993). Presumably that is why these demographic variables were not included in their model predicting depressive symptoms.

Since these demographic factors are standard confounders for health measures, it is perhaps more surprising that there were studies that omitted some measure of SES, particularly the NHEFS. Comoni-Huntley, Barbano, Brody, Cohen, Feldman, Kleinman et al. (1983) detailed the information obtained for the NHEFS and it included standard epidemiologic demographic variables, as one would have expected. A rationale for why these variables were not included in the analysis by Busch et al. (1994) was not given.

### Health Status

Current health status as well as history of depression are well documented as predictors of current depressive symptoms (Dennerstein et al., 1994; Kaufert et al., 1992;

McKinlay et al., 1987). Woods & Mitchell (1996), Avis et al. (1994), Kaufert et al. (1992), and Hunter (1990, 1992) included measures of health status in their analyses. Woods & Mitchell (1996) used a variable that resulted from factor analysis of many measures. Avis et al. (1994) used a single combined variable of hot flashes, night sweats, and menstrual problems, as discussed under the section on vasomotor symptoms. Kaufert et al. (1992) used three measures, a three-valence rating of present health, presence or absence of four of the most common diagnoses from their cross-sectional study, and a total sum of current diagnoses. Hunter (1992) used a variable that resulted from factor analysis, similar in approach to Woods & Mitchell (1996). Because of the analytic strategies used, Avis et al. (1994), Bromberger & Matthews (1996b), and Hunter (1990, 1992) were able to include previous depression as an independent predictor of the outcome of later depressive symptoms.

#### Attitudes and Cultural Beliefs

Cultural beliefs about menopause were found to be significantly predictive of depressive symptoms by cross-cultural researchers (Kaufert et al., 1986). For example, those cultures, such as ours, that believe menopause is bad for women also believe that depression must accompany it. Nevertheless, only Woods & Mitchell (1996) and Hunter (1990) measured this variable. Hunter (1990) and Holte (1992) both included a measure of expectations of menopause, separate from beliefs, as well.

#### Hypothalamic-Pituitary-Ovarian (HPO) Hormones

An extensive literature has explored theories that hypothalamic-pituitary-ovarian hormones are related to depressive symptoms (Nolen-Hoeksema, 1990, Chap. 3). The

HPO hormones that are of interest during the menopausal transition include estradiol (E2), estrone (E1), follicle-stimulating hormone (FSH), and testosterone (an androgen). While no consistent evidence has supported such theories, the underlying belief that produced such theories pervades much medical research on depression (Nolen-Hoeksema, 1990, Chap. 3). Current studies, such as the SMWHS and the Study of Women's Health Across the Nation are underway to investigate relationships between changing hypothalamic-pituitary-ovarian (HPO) hormone profiles in perimenopausal and postmenopausal women and depressive symptoms. So far, only the Melbourne Women's Midlife Health Project has included hormonal assays in published analyses of longitudinal depressive symptoms (Dennerstein et al., 1997).

In summary, the diversity in these common (and presumably, therefore, important) other independent variables made comparisons across studies problematic. Diversity occurred at both conceptual and operational levels: for some researchers variables here termed other independent ones were considered as outcome or dependent variables with the same status as depressive symptoms. For example, Holte included vasomotor symptoms as a separate outcome variable, not a predictor of depression. Several studies found correlations between depressive symptoms and other symptoms, often explained by personality type; women who were depressed reported more symptoms in general, including vasomotor symptoms, stress, and poor health status, and/or were likely to have low SES (Avis et al., 1994; Bromberger & Matthews, 1996b; Hällström & Samuelsson, 1985; Hunter, 1992; Kaufert et al., 1992; Woods & Mitchell, 1996).

### Designs

All of the studies reviewed here were prospective and observational in design. Longitudinal research designs must also specify the number and timing of data collection points (see Table 3). Three of the nine longitudinal studies measured depressive symptoms more than two times: Kaufert et al. (1992) measured six although they had to discard the final one in analysis because of problems with implementation. Dennerstein et al. (1997) measured four, and Holte (1992) measured six although the latter aggregated results into two summary measures. None of these were U.S. studies.

Timing of the data collection ranged from six months apart in Manitoba to 10 years apart for the NHEFS, with three studies using a time interval of one year. While Avis et al. (1994) contacted Massachusetts women every nine months, the CES-D was administered at only two time points 27 months apart. Those studies that used the most number of measurement occasions also used the shortest time intervals (see Table 3). No researcher provided a rationale for the time interval either in terms of theory or existing research regarding changes in depressive symptoms or changes in menopausal status. Research on MDD has indicated that several months are required for intervention to have effect for those women that are diagnosed, treated, and respond (Depression Guideline Panel, 1993; Evans et al., 1992; Frank, Grochocinski, Spanier, Buysse, Cherry, Houck et al., 2000; Kupfer et al., 1992; Miller & Keitner, 1996; Shea et al, 1992; Simon, Katon, Rutter, Von Korff, Lin, Robinson et al., 1998). Therefore, measuring CES-D scores every six months, as the Manitoba researchers did, would be very likely to capture change from an episode to remission or recovery, and measurement every nine or even 12

months, as the Massachusetts, Seattle, Melbourne, and Norwegian researchers did, would also be likely to identify changes in depressive symptoms. Longer time intervals between measurements could possibly miss depressive episodes and/or recoveries. As for MTS, middle and late transition combined have been determined to last approximately 6 years on average (Ellen S. Mitchell, personal communication). Thus, depressive symptoms would have to be measured over several years in order to allow enough of the women being followed to change at least once from one MTS to another for true longitudinal design.

Busch et al. (1994) did discuss how the 10 years between their two measurements points might have, in fact, missed any changes in depressive symptoms that occurred during the rather long unmeasured time interval. Woods & Mitchell (1996), who used only two time points one year apart, discussed that this time interval was possibly too short to allow for much expected change in MTS. By design Holte (1992) could not identify changes related to menopausal transition, only to the one time point of menopause itself.

Longitudinal studies must balance breadth of variables at any one time point with length of the study, and with burden on the participants. The number and type of variables might be limited depending upon the cost and feasibility of measuring them repeatedly, as well as on the burden for participants (Avis et al. 1994) and the possibility of creating a bias from familiarity with the questions. There must also be a balance with sample size. Telephone interviews and especially questionnaires are relatively easy and inexpensive. It cannot be considered a coincidence that the largest study, Avis et al.

(1994), was entirely conducted by telephone interview. It is remarkable that the Gothenburg study of 800 women included in-person expert interviews up to two hours each and required women to spend at least a half day further for a battery of physical and physiological measurements (Samuelsson, 1982).

### Analysis Strategies And Techniques

Most of the longitudinal studies of transition to menopause to date used fundamentally different analytic strategies to investigate the outcome of depressive symptoms. The first group of studies reviewed here were those that used strategies consistent with statistical assumptions, only one of which used multiple time points longitudinally. The next group reviewed were those that, while having multiple time points, used analytic techniques that failed to accommodate inter-individual correlation of repeated measures.

### Norwegian Menopause Project

Of those using analytic strategies consistent with statistical assumptions, only Holte (1992) measured depressive symptoms at more than two time points. As Lavori (1990) pointed out, comparing summary measures from before and after an event are a statistically appropriate method to test for effects of the event. Holte (1992) identified menopause as the event and mean depression scores pre and post menopause were computed. A t-test for paired samples was performed to compare pre and post summary measures, the mean depression scores. However, this strategy did, in effect, reduce the data set to two measurements, thereby losing some available information.

While statistically appropriate, this strategy could not provide information about changes occurring during the transition to menopause or during the early postmenopause. It apparently assumed effects were related to the singular moment of menopause itself, an assumption somewhat belied by the individual graphs Holte himself provided (p. 132). By design, therefore, any changes in depressive symptoms associated with the menopausal transition itself were obscured by using a mean premenopause score. The problem of losing information through aggregating individual measurements into summary measures has been discussed by Bryk & Raudenbush (1992, Chap. 1) and Heck & Thomas (2000, Chap. 1).

Other criticisms of the Norwegian Menopause Project included 1) the definition of menopause, which was an atypical one (see section on Menopausal and Menstrual Status earlier); 2) the exclusion of those women who had bleeding after six months of amenorrhea, likely to have been women in what would have been categorized as prolonged perimenopause in the Massachusetts study; and 3) exclusion of women on HRT and/or with surgical menopause. Because of these exclusions the sample size ended up being small in the longitudinal portion of this study, 59 from an original 200 women randomly selected from the cross-sectional study sample of 1668.

Holte performed comparisons on 87 single measures covering 26 areas of health complaints. He appropriately used paired t-tests for those single measures that were normally distributed and Wilcoxon's matched-pairs signed-rank tests for those with bimodal or skewed distributions. While differences in 14 single measures in six areas of health complaints were found to be statistically significant (not including depressive

symptoms), there was no mention of what, if any, mechanism was used to deal with the problem of the 87 multiple comparisons (Fisher & Belle, 1993, Chapter 12). By chance, several though not so many as 14, of the 87 measures would have been expected to be statistically significant at a 95% level of confidence, although the significance level was not given. Thus, while the analysis strategy was appropriate, the rationale for which was well discussed in the article, several other aspects of analysis and its reporting were less than optimum.

#### Seattle Midlife Women's Health Study

In the Seattle study, Woods and Mitchell (1996) created four categories of the outcome variable—continuous absence of depression, consistent depression, recovering from depression, and emerging depression—allowing for discriminant analytic techniques. While this strategy was a useful one for a study with only two time points, it cannot be easily extended to multiple time points since the number of outcome categories increases by an exponent of two with each additional time point. In fact, an attempt to analyze more than two time points with the same strategy was attempted and abandoned as not feasible, according to a Co-Investigator (Ellen Mitchell, personal communication).

#### University Of Pittsburgh Healthy Women Study

Bromberger and Matthews (1996b) used “hierarchical process for linear regression,” or forced-entry stepwise multivariate regression to predict Time 2 depressive symptoms. Women were required to be premenopausal at entry, and three groups were created at Time 2, women still premenopausal, a combined peri- and postmenopause group, and women on HRT. Their use of combined peri- and postmenopause MTS has

already been reviewed. They included a score of depressive symptoms at Time 1 in the first step of the regression, after excluding women from the longitudinal study who had a Beck Depression Inventory score of  $>9$  at Time 1. Various psychological variables were included in the model as predictors as well as MTS. While this approach was appropriate for a longitudinal study with only two time points, it cannot be easily extended for multiple time points.

#### Massachusetts Women's Health Study

Avis et al. (1994) used multiple logistic regression models, testing them for goodness of fit with the data. They calculated odds ratios for predictors and covariates of Time 2 depression, a dichotomized score of depressive symptoms. Included in the models as predictors were Time 1 depression, HRT use, change in menopausal status between Times 1 and 2, and in a second model menopausal symptoms at Time 2, as a dichotomous variable. The same comment as in the section on the University of Pittsburgh study about extending logistic regression to multiple repeated time points applies.

Like Kaufert et al. (1992), Avis and colleagues looked at change in MTS. The Massachusetts researchers coded the changes in MTS as a set of indicator variables: premenopausal at both time points, which was the reference category; perimenopausal at both time points; postmenopausal at both time points; changing from pre- to perimenopausal; and changing from either pre- or perimenopausal to postmenopausal. The rationale for combining both pre- and perimenopause at time one for this category

was not provided; it might have been related to number of women in this subcategory or it might have been based on a model that attributed change to menopause itself.

#### National Health Examination Follow-Up Study (NHEFS)

The NHEFS used repeated measures analysis of variance, or MANOVA. They succumbed to the criticisms of Ekstrom, Quade, and Golden (1990) and Lavori (1990) by not providing enough information for the reader to know the details of how MANOVA was performed in order to evaluate the use of the analytic method. They reported merely that “repeated measures analyses of variance” were used. Degrees of freedom of the F statistics were not reported. No mention was made of adjustment for heterogeneous correlations among the repeated measurements. Heterogeneous correlations refers to the fact that the two depression scores for the same woman would be expected to be more correlated than depression scores between different women.

While adjustment may have been done and not reported, another of the assumptions of MANOVA might (also) have been violated, that of normal distribution of each of the variables at each time point of measurement. Again, no information was provided by the authors as to whether a transformation of the CES-D score was used in order to create a variable with a more normal distribution. Since CES-D scores are not expected to be normally distributed (Avis et al., 1994, p. 215; Devins & Orme, 1985; Sheehan, Fifield, Reisine, & Tennen, 1995) such a transformation would have to be performed in order to avoid violating MANOVA assumptions.

### Longitudinal Study Of Women In Gothenburg

From Hällström & Samuelsson (1985) scant information was provided about statistical methods used. The Swedish researchers listed their statistical analytic methods in a preliminary paragraph without describing the tests, apparently assuming any reader would be familiar with them. All tests were non-parametric—Pitman's permutation test, Fisher's permutation test, or Fisher's exact test, although which method was used for which result was not further specified. Results were presented as number of women in a specified category, percentages of women in a specified category, or rates of onset of mental health problems per year. For many of the results, significance is reported only in the narrative, not in the charts.

The issue of multiple comparisons was not addressed even though their data were analyzed from several different approaches. Lastly, the validity of using recalled information for a retrospective diagnosis was not addressed. In general, not enough information was provided by Hällström and Samuelsson to evaluate the appropriateness of their use of statistical analyses.

### South-East England Study

Hunter used paired t-tests and McNemar tests to determine changes between the first and second measurement points in depressive symptoms. Multiple stepwise regression was also used as an analysis method to predict symptoms at Time 2, regressed on Time 1 measurement of depressive symptoms, premenopausal vasomotor symptoms, history of premenstrual tension, negative beliefs about menopause, employment, and social class.

### Manitoba Project On Women And Their Health In The Middle Years

In order to measure the concept of change the Manitoba researchers considered their data in pairs of measurements, as the change (or lack of change) between two consecutive time points. Thus they considered each woman's Time 1 and Time 2 measures as a pair, each woman's Time 2 and Time 3 measures as a pair, each woman's Time 3 and Time 4 measures as a pair, and each woman's Time 4 and Time 5 measures as a pair. In their longitudinal study of menstrual status change (Kaufert et al., 1987) where they used the same approach, they explained that

The statistical acceptability of collapsing the data into a single table [of Time 1 and Time 2 measurements] depends on whether patterns of menstrual change differ from one table to the next. This approach to the analysis of longitudinal data amounts to "treating each interval between events for each individual as a separate observation<sup>8</sup>."

No reference 8 was included in this published article, so the authority of this source could not be evaluated. They were appropriately concerned about the validity of this step, so in their study of depressive symptoms (Kaufert et al., 1992) they analyzed correlations between Times 1 and 2, between Times 2 and 3, 3 and 4, and 4 and 5 from the entire data set.

The appropriateness of this procedure [collapsing the repeated data of five time points into pairs of time points] depends on the stability of patterns of change from one interview to the next over the three years of the study. The hypothesis that the transition probabilities were constant over time was tested using the Mantel-Haenszel chi-square, first for menopausal status ( $\chi^2 = 25.1$ ,  $df = 24$ ,  $P = 0.39$ ) and then for depression ( $\chi^2 = 5.9$ ,  $df = 6$ ,  $P = 0.43$ ). These results suggest that the probability of changing from one state to another was independent of the time interval [presumably all time intervals, from 1 to 2, 2 to 3, 3 to 4, and 4 to 5], and that we might collapse the data into a single table (p. 149).

Their concern about the appropriateness of pooling (their term) the data was laudable. However, pooling pairs of measurements; that is, considering each pair as independent of each other, still ignored the issue of correlation within woman between *all* time points, for example between Time 1 and 3, 1 and 4, 1 and 5, 2 and 4, 2 and 5, and 3 and 5, and between the first pair (Time 1 and Time 2), the second pair (Time 2 and Time 3), the third pair (Time 3 and Time 4), and the fourth pair (Time 4 and Time 5) from the same woman. The expectation of longitudinal data is that the repeated measures on the same woman are not independent but correlated, more closely between consecutive pairs of measurements and less closely as the pairs are less contiguous. The effect of not accommodating correlation is to underestimate the standard error thus overestimating significance (Heck & Thomas, 2000, p. 1). The underestimated standard error is in the denominator of a significance statistic, such as the t- or F- or  $\chi^2$  statistic. Thus it will make the significance statistic too big. The probability will therefore be too small, finding that there is a significant difference when in fact there is none.

The Manitoba researchers first used a  $\chi^2$  approach for analysis of menopausal status change and depression. The advantages of the  $\chi^2$  analyses might have been their general ease of use, recognizability and acceptance. Two by two contingency tables were created for depression/ non-depression at time one and depression/non-depression at time two. However, because of the limitations of the contingency table  $\chi^2$  analyses Kaufert and colleagues were required to run separate analyses for each predictor such as MTS. In turn, the MTS had to be dichotomized as pre vs peri plus post, peri vs pre plus post, and

post vs pre plus peri and separate analyses conducted for each dichotomy. After performing these numerous  $\chi^2$  calculations without mentioning accommodation for multiple comparisons, they developed logistic regression models using a backward stepwise fitting algorithm, one model for women not depressed at “Time 1” and a separate one for women depressed at “Time 1.” Again, assumptions of ordinary least squares regression that each measurement, here pairs, be independent (Asher, 1983, p. 25-30; Kleinbaum et al., 1998, p. 107) were not accommodated.

#### Melbourne Women’s Midlife Health Project

The Melbourne researchers also collapsed their data set so that repeated measures were handled as though they were independently obtained from different women. Instead of four repeated measures from 400 women they also pooled (their term) their data as 1600 independent data points. Thus, their analysis strategy transformed this study into a cross-sectional one. These researchers were therefore unable to determine any effects of change over time, either in menopausal status or in depressive symptoms.

Similar to the Manitoba study, the Melbourne study ignored the correlated nature of longitudinally repeated measures and thus failed to observe assumptions of logistic regression about independence of observations. Dennerstein et al. (1997), themselves, identified their problem of having used a “generalized linear model assuming normal errors” (p. 97):

The assumption of independence of observations within individuals was examined by calculating the between-year correlation of the residuals from the fitted models. For the modeling of positive and negative affect, these correlations ranged from 0.43 to 0.65 and 0.32 to 0.56 respectively.

**These correlations indicate that the regression estimates should be interpreted with caution (p. 100-101).**

**Furthermore, their strategy of repeatedly running different regression models until coefficients changed, in order to establish a final model, has been criticized by a research consultant as less than ideal for interpretation of results (Martha Lentz, personal communication). Such a technique might be described as exploratory rather than confirmatory in nature. It is preferable to develop theory- or conceptually-based models for testing so that results can be interpreted as consistent with or not consistent with theory or concepts.**

**In summary, two of the three longitudinal analyses of multiple time points used analytic strategies that could find statistical significance where there was actually none by ignoring correlations between repeated measures on the same woman. They also lost the advantage of collecting repeated measures longitudinally on the same individual. Additionally, the Melbourne study lost the ability to investigate change. As Holte (1992) and Lavori (1990) have pointed out, why go to the considerable effort and expense of tracking the same individuals and repeating measurements if you then lose the ability “to increase the precision of comparison by reducing the statistical variability of measurement’ (Lavori, 1990, p. 776)?**

**One answer may be the relative obscurity and difficulty of longitudinal analytic methods, as well as their recent appearance for research. Analysis strategies for multi-point longitudinal data patterns have not been taught commonly in statistics classes in the past, and the underlying statistics have been complex and somewhat difficult to find in**

published form (Collins & Horn, 1991; Ekstrom et al., 1990). As late as 1990 Lavori discussed researchers' interest in "detailed temporal patterns of difference" (p. 777). He added, "in my experience, when this kind of pattern analysis is discussed, everyone gets very excited; eyes light up at the prospect of fascinating and biologically informative complex differences in profiles" (p. 777). He ended without mentioning any new or improved method for fulfilling this interest and excitement, however, while critiquing RM-ANOVA. The American Psychological Association conference whose proceedings were published as Best Methods for the Analysis of Change (Collins & Horn, 1991) attempted to fill the void. Unfortunately, many of the methods required creation of statistical applications by the individual authors and were not generally available even if knowledge of them were. Structural equation modeling packages were also difficult to find for a few years due to marketing changes, according to a statistical consultant (Robert Burr, personal communication, 1998).

New techniques and software are becoming more available recently, however. Both structural equation modeling (Duncan-Jones et al., 1990) and hierarchical linear modeling (Bryk & Raudenbush, 1992, Chap. 6) are statistical techniques that have been used for longitudinal analyses of change over time. Plans to conduct analyses along these lines were mentioned by Dennerstein and colleagues (1997) and are also planned by the team of researchers conducting the SMWHS.

Kaufert et al. (1992) have discussed the importance of the concept of change over time. In longitudinal studies this concept, along with that of causation, would seem to be key. If a researcher hypothesizes that being in a particular stage is associated with a

particular outcome, then longitudinal design is not required. If, however, the hypothesis is that being in a particular stage causes a particular outcome, longitudinal strategies would be advantageous. In addition, a hypothesis that change causes a particular outcome, or some quality of the stage such as its being prolonged is causal, longitudinal strategies would be required to test such an hypothesis.

### Results

Given the difficulties already reviewed in comparing these nine longitudinal studies of depression during the menopausal transition, the general consistency of negative results is all that more remarkable. Five of the studies reviewed here found no relationship between menopausal status or a change in menopausal status and depressive symptoms. All of the studies that were designed for it found that the single best predictor of depression at a later time point was depression at an earlier time point, unrelated to menopausal status or its change. Three studies found significant relationships between depressive symptoms and menopausal status: South-East England, the only study to use clinical sampling; Massachusetts (1994); and Melbourne (1997), the last being one of the two studies to use analysis techniques that can produce statistical significance falsely. First, each study's results have been summarized for the relationship between depression and menopausal status. Secondly, effects of covariates or predictors are reviewed. Lastly, comparisons to other studies have been included where meaningful.

#### Seattle Midlife Women's Health Study

Woods and Mitchell did not find menstrual changes to significantly predict depressive symptoms or their pattern of change. However, the Seattle researchers did

find different variable groups from a factor analysis to discriminate between change patterns, including previous history of depression such as premenstrual syndrome or postpartum depression, stress, health status, socialization, family resources and social support. The pattern of consistent depressed mood was discriminated from the pattern of resolving depressed mood additionally by vasomotor symptoms. The predictors could be interpreted as history of previous depression and life events and stressors, allowing for a buffering effect of resources and support. As an additional note about pattern, while these authors did not provide the information in this form, 22% of women changed category of depressive symptoms in one year, 41 of 347 increasing their scores to over the CES-D cutoff of  $\geq 16$ , and 36 of 347 decreasing below it. Thus, 78% of the women were stable over the one year interval.

#### University Of Pittsburgh Healthy Women Study

Bromberger & Matthews (1996b) found that neither HRT use compared to premenopausal status, nor combined peri- and postmenopausal status compared to premenopausal status, contributed significantly to the outcome of depressive symptoms. They also examined individual correlations between each individual measure and depressive symptoms at follow-up and found that depressive symptom score at entry had the highest, and statistically significant, correlation. Depressive symptom score at Time 1 was also the significant predictor of highest magnitude in a multiple regression model for prediction of Time 2 depressive symptoms. Measures of trait anxiety, stress as an ongoing problem, and stress as a recent event were also significant predictors at lesser

magnitudes. Thus, life events and stressors were also predictive of midlife depressive symptoms.

### Massachusetts Women's Health Study

Of the three studies where significant differences in depressive symptoms by menopausal status were obtained, the Massachusetts study (Avis et al., 1994) was the one with the largest sample size. They found significantly higher depression scores in the group of women who were perimenopausal at both time points, 27 months (2.25 years) apart. This higher rate of depressive symptoms was found when compared to women who were premenopausal at both time points, women who were postmenopausal at both time points, and women who had changed status. The measure of significance was not dramatic, with the p value given as  $< 0.05$  and the lower limit of the 95% confidence interval of the odds ratio given as 1.05; a confidence interval including 1.00 would not have been significant. The significant difference between the group of women perimenopausal at both time points and other groups was not maintained when the logistic regression model was adjusted for menstrual problems. The significant results of the Massachusetts study can be evaluated in terms of the sample size, the definition of menstrual problems, and the definitions of pre- and perimenopause.

The Massachusetts study had by far the largest sample of the nine studies reviewed in this chapter, which makes their results noteworthy but also enhanced their chances of finding statistical significance for small differences. Hot flashes, night sweats, and "menstrual problems" were combined into a single predictor variable in their logistic regression model. The reason these variables were combined was because "none

of the women in the pre/peri-post and post-post groups reported menstrual problems” (p. 216). As explained in the section on other independent variables, “menstrual problems” (Magursky et al., 1975) were understood to be the defining criteria of early and middle transition. It was therefore not surprising that it was mostly the peri-peri group that complained of these “problems.” However, when the model was adjusted for menstrual problems alone without vasomotor symptoms, there were still no differences in depression scores between MTS groups.

In their discussion Avis et al. (1994) acknowledged that menopausal symptoms might be an expression of depression; in that case they would have adjusted for a dimension of their outcome variable itself, an adjustment that did indeed result in loss of significance between MTS groups. In other analyses the Massachusetts researchers noted that women who had psychological complaints also tended to have complaints about all symptoms (McKinlay et al., 1992) and to have had most of the same complaints prior to menopausal transition (Avis, Crawford, & McKinlay, 1997). Women were excluded in the longitudinal analysis who had had hysterectomies by the second time point, a group that had the highest CES-D scores in their cross-sectional analysis (McKinlay et al., 1987).

Definitions of perimenopause and prolonged perimenopause might also have been problematic. As detailed earlier, their definitions of MTS were not mutually exclusive. Women could have experienced three months of regular periods immediately before the interview and still have experienced periods of amenorrhea and/or changes in regularity or flow during the previous 12 months. Since women who were complaining of

menstrual problems, that is, complaining of the characteristics that define middle and late transition, were also the same women who tended to complain about all symptoms, including depressive ones, it was possible that women who were classified as perimenopausal, especially at two time points, were those women who were more likely to have depressive symptoms. This misclassification bias would explain the significant results of prolonged perimenopause being a predictor of high CES-D score.

The Massachusetts researchers also labeled the pattern of repeated perimenopausal status at two measurements as “prolonged” perimenopause even though it lasted only 27 months (2.25 years). The implication was that in the Massachusetts study there was something unusual about perimenopause lasting 27 months. However, it was these same researchers who provided the basis of the WHO claim that “the average duration of the menopausal transition is 3.8 years” (McKinlay et al., 1992; *Research on the Menopause*, 1996, p. 14). The presentation of a 27-month perimenopause as “prolonged” might have been misleading.

Avis et al. (1994) described their results as a transitory increase in depressive symptoms during the perimenopause. In actuality, however, it was not known if the persistently perimenopausal women had low depression scores before perimenopause and/or would have low scores when postmenopausal. A transitory increase does not explain why women who were perimenopausal at the second time point only, premenopausal becoming perimenopausal, did not show a significant difference from the reference group of women premenopausal at both time points. Their ability to determine if depression scores decreased after perimenopause was also limited by not using a

change category of peri becoming post. They combined premenopause and perimenopause into one for a change category of pre/peri becoming post. Similar to the University of Pittsburgh, Manitoba, and Seattle studies and an earlier cross-sectional analysis (Avis et al., 1997), their longitudinal analysis found a previous history of depressive symptoms to be the strongest predictor of perimenopausal depressive symptoms.

Because of the differences in designs and analyses between the Massachusetts study and the other studies reviewed here, however, neither refutation nor replication of the Massachusetts results were possible. The University of Pittsburgh study combined perimenopausal and postmenopausal women into one category and thus by design could not have determined an effect of any form of perimenopause alone, prolonged or otherwise, only the effect of change to not having menstrual periods within at least the previous three months. The South-East England study admitted women to the longitudinal study only if they were premenopausal at the first time point and combined perimenopause with postmenopause; similar to the University of Pittsburgh study, by design the South-East England study could not have determined effects of any form of perimenopause, prolonged or otherwise.

Woods & Mitchell (1996) did not include a long enough time span, only 12 months apart, to be able to replicate the prolonged perimenopause of the Massachusetts study. Also, the majority of the Seattle women were in early or middle transition, not late transition when MTS changes would have been most similar to “prolonged perimenopause.” The NHEFS and Gothenburg study measured depression at time points

too far apart, 10 years and six years respectively, to replicate any effects of “prolonged perimenopause.” The NHEFS pointed out themselves that significant increases in depression scores might have occurred and resolved within the 10 years between their two data collection points. Also, by design, all women in the NHEFS were premenopausal at time one, so for that reason, too, they could not replicate the findings for prolonged perimenopause of the Massachusetts study. Again by design, Holte did not include perimenopause as a variable and thus could not have determined any effect it might have had on depressive symptoms.

The Manitoba study (Kaufert et al., 1992) analyzed change only between time points six months apart and could not have replicated any effects of prolonged perimenopause either. It was possible that if Kaufert and colleagues had examined data for women who remained perimenopausal across 27 months or longer and then compared them to other women they might have replicated McKinlay’s findings. The Melbourne study also “pooled” their data set so that each datum point was analyzed as if independent, so they, too, could not have found any effects of “prolonged perimenopause.”

### NHEFS

Busch et al. (1994) found no effect of MTS change group on depressive symptoms, and otherwise measured only age within MTS. They included a caution that in their study “subjects exceeding the cutoff point for depression [at Time 1] were less likely to be traced [for Time 2 follow-up] than were non-depressed subjects” (p. 225), possibly affecting the negative results through selection bias. Because of the 10-year

time interval, the lack of covariates or predictors, and the recognized bias in the follow-up sample, it is difficult to compare these results to those of other studies.

**Manitoba Project On Women And Their Health In The Middle Years**

The Manitoba study found no relationship between change in depression score and menopausal status or change in menopausal status and depression score. Starting with  $\chi^2$  analyses these researchers did find significantly increased likelihood of depression scores at “time two” in those women with hysterectomy (at any time) who had not been depressed at “time one,” compared to all other women. The relative odds of depression at time two in this group were 1.7 (95% C.I. 1.15-2.6).

The multiple logistic regression models with backward stepwise entry of variables confirmed the results of the series of  $\chi^2$  analyses where “menopausal status categories were entered as dummy variables” (p. 153). After the other independent variables were adjusted for, variables that were predictive of higher depressive symptoms at time two after lower depressive symptoms at time one included poor self-rating of health status, diagnoses of arthritis or high blood pressure, self rating of problems with husband, children, other relationships, or other areas of life. Variables predictive of staying depressed at time two after being depressed at time one again included a poor self-rating of health status and diagnoses of three or more other health problems. Confusingly, diagnosis of arthritis or thyroid problems was reported in the narrative results as significantly associated with staying depressed; however, significance levels were reported as .08 and .04, respectively, but the confidence intervals reported for both of these odds ratios included 1.0.

Understanding the effects of hysterectomy were also problematic. Because of the technique of pooling or collapsing the longitudinal data into pairs of time points, “time one” might have been any one of the measurements 1 through 4; “time two” might have been any one of the measurements 2 through 5. Hysterectomy might not have occurred during the particular six-month interval between “time one” and “time two” when rise in depression score occurred. Yet if hysterectomy had occurred at any time during the study, the woman was more likely to have a pair of six month depression scores where score at “time two” was high after a high “time one” score. As discussed already, the analytic technique of pooling made significant results uninterpretable. Another possible explanation for the Manitoba study’s results was a Type I error resulting from “a series of”  $\chi^2$  analyses performed (p. 150) without accommodation for multiple comparisons; this one significant result was due to chance.

Comparison of these significant results with other studies were not possible because most of the studies excluded women who had hysterectomies; only the Massachusetts, NHEFS, and the University of Pittsburgh studies included women who had hysterectomies, and the University of Pittsburgh study included them in the HRT group in general. These other studies found no evidence of increased depression scores in the HRT/hysterectomy group. Thus, of the four studies that included women with hysterectomies, only Kaufert et al. (1992) found hysterectomy to be predictive for increase in depression scores, and then only in the subgroup who were not consistently depressed. Of the other three studies that found significant increase in depressive

symptoms with some stage of MTS (Avis et al., 1994; Hunter, 1990), none included women who had had hysterectomies.

Another set of results from the Manitoba study were unique, however. This study was the only one, other than a few case examples in the Norwegian article (Holte, 1992), that presented results of longitudinally repeated measures, although no statistical analyses were performed. While Kaufert et al. (1992) were critiqued in this review for analytic strategies that failed to account for inter-individual correlation of repeated measures, this presentation avoided that criticism. They displayed patterns over time of CES-D scores along with the number of women demonstrating each pattern. The CES-D score was classified as dichotomous,  $\geq 16$  and  $< 16$ , in black and white boxes, respectively. Even with some patterns not present, they observed 29 out of a possible total of 32 or  $2^5$  patterns (p. 148). While numbers of women demonstrating each longitudinal pattern were provided, little interpretation or discussion was added. In their earlier article using the same format for presentation of longitudinal patterns of menstrual changes, they wrote that “this manner of presentation is too detailed for interpretation” (Kaufert et al., 1987, p. 221). These intriguing results of longitudinal patterns have been compared in detail to the current study’s results in Chapter 5.

#### South-East England Study

Hunter (1990, 1992) was the only researcher who found in both cross-sectional (Hunter, Battersby, & Whitehead, 1986) and longitudinal analyses that the group of combined peri- and postmenopausal women experienced increased depressive symptoms over premenopausal women. These results might have been due to the combining of

and/or definitions of MTS, cultural differences of women in South-East England, the small sample size (36 total), the use of clinical sampling, or chance results. As already explained, the use of the clinical sample was the most likely reason for the significance of the results (Kaufert & Syrotuik, 1981). Bromberger & Matthews (1996b) also combined peri- and postmenopause and compared depressive symptoms to premenopause but found no significant differences. Results were not comparable to other studies because of the sampling methods and use of MTS.

#### Melbourne Women's Midlife Health Project

Dennerstein and colleagues found significant differences in depressive symptoms comparing premenopause to perimenopause and comparing premenopause to early postmenopause only in their longitudinal analysis (Dennerstein et al., 1997), not in their cross-sectional one (Dennerstein et al., 1993). In the longitudinal study, negative affect in late postmenopause also was not significantly different from premenopause. Higher negative affect for peri- and early postmenopause remained significant even after hot flashes were included in the regression model, although this finding was not true for a measure of positive affect. None of the hormonal assays of estradiol ( $E_2$ ), free androgen index (FAI), follicle stimulating hormone (FSH), or serum hormone binding globulin (SHBG) or hot flashes contributed significantly to the model of CES-D score after MTS was already entered. Since hormonal assays and MTS had been found to be correlated, this result was not surprising.

Similar to Avis et al. (1994), the Melbourne study described the increase in negative affect as temporary because of the lack of significant difference between

premenopause and late postmenopause. However, in the Melbourne study, changes in depression scores across time for an individual woman could not be determined because of their analysis strategy of treating longitudinal data as a cross-sectional data set.

By their own admission this analysis did not include the variables of health status, stress, marital status, premenstrual complaints, and lifestyle that these same researchers found “significantly related” to well-being (Dennerstein et al, 1997, p. 101) in their cross sectional analyses (Dennerstein et al., 1994; Dennerstein et al., 1993). In light of the inter-individual correlations in the “pooled” longitudinal data set and the omission of previously significant predictors in the model, a finding of statistical significance was uninterpretable.

#### Norwegian Menopause Project

The only study using multiple time points and an appropriate analytic strategy did not find any relationship between the menopause and depressive symptoms. Social dysfunction, a possible indicator of depression, was significantly related to vasomotor complaints. Difficulties in comparing this study with others based on the analytic strategy and design have already been reviewed.

#### Longitudinal Study Of Women In Gothenburg

The Gothenburg researchers were the only ones to diagnose depressive disorder, although only as one category of mental health problems albeit the majority of them. They also measured level of disability from mental health problems. They found no relationship between mental health problems or disability and MTS. These researchers also found, by retrospective recall, that the best predictor of a major depressive episode in

midlife was a history of earlier depressive symptoms. The Gothenburg study also investigated onset rate or incidence of mental disorders, using recall for the previous year, and found that rates were constant and also unrelated to MTS. The implications of this finding were that some women did experience a first major depressive episode during midlife, but the rates for first episode were not increased because of menopausal transition. Women who were divorced, childless, and of lower SES had higher prevalences and incidences of mental disorder than women of higher SES who were married with children. These predictors could be interpreted as life events and stressors.

While not statistically significant, a trend toward a cohort effect was evident. Women who were in the oldest age group at entry had higher prevalences of mental disorders at both measurement points than younger women at either measurement point. There was no trend for increasing prevalence within any age cohort from first to second measurement point, however, regardless of MTS. Because the Gothenburg study was the earliest longitudinal research on depressive symptoms and menopausal transition (first wave in 1968), its findings of no relationship between these two phenomena were reassuring about the lack of cohort effects.

### Summary

In sum, few longitudinal studies of depression or depressive symptoms in midlife women have found increase in depression, and those few used clinical sampling (Hunter, 1990, 1992), or statistical methods whose assumptions were not met (Dennerstein et al., 1997; Kaufert et al., 1992), or definitions of MTS that were not mutually exclusive and might have biased inclusion of women with higher depressive symptoms into the

perimenopause category (Avis et al., 1994). In other words, the longitudinal studies of depression during the menopausal transition that found an increase in or higher depression during perimenopause had methodological flaws to which the differences can be attributed. The majority of longitudinal studies found no relationship between depressive symptoms and menopausal transition or menopause.

Those studies that were able to include it in their models or analysis strategies found that the single best predictor of depressive symptoms at the second time point was level of depressive symptoms at the first. This finding was consistent with those of the course of depression; it would be expected that by midlife a majority of depressive episodes would be recurrences or relapses, perhaps not any more likely to be related to menopausal transition than recurrences or relapses at other ages would be related to menses (Winokur, 1973).

Those studies that were designed to include a measure of life events or stressors (Bromberger & Matthews, 1996b; Hällström & Samuelsson, 1985; Hunter, 1992; Kaufert et al., 1992; Woods & Mitchell, 1996) found significant relationships between some measure of life events and/or stressors, although Bromberger and Matthews found significance only in an interaction with anxiety and Hunter only when previous depressive symptoms were removed from the multiple regression model.

General criticisms of the nine studies reviewed here included sampling frames and methods that limited external validity (South-East England); sample sizes that were too small to establish reliability of estimates (South-East England, Norway); a general assumption of comparability across different countries; lack of consistency in definition

of an important independent variable, MTS; and except for three studies, measurement at only two time points. Often, studies excluded an important segment of the sample related to the outcome variable; several excluded HRT users, women with surgical menopause, or, as the Norwegian study did, excluded women in what might have been “prolonged perimenopause.”

Six of the studies were limited to only two time points, in one study very close together (one year) and in one very far apart (10 years). Only one of the three studies with multiple time points of measurement used an analytic technique that accounted for inter-individual correlation, and that study (Holte) aggregated observations into two summary measures for a simple comparison pre and post FMP, thereby losing some longitudinal information about patterns. Two of the three studies with multiple time points used analyses with “pooling” of repeated measures as if they were independent observations.

Each of the studies reviewed have contributed to understanding changes in depressive symptoms during the menopausal transition and how best to study them. Woods & Mitchell (1996) investigated the rich context of women’s lives, including resources and support, for analysis of depressive symptoms; they also investigated change during midlife. Avis et al. (1994) utilized the largest sample size, thus providing the most reliable estimates of effects, and included a measure of change. Bromberger and Matthews (1996) included consideration of psychological theories of depression, in particular theories relating depression to personality, and included a measure of change. Busch and colleagues (1994) included the epidemiologically important categories of

surgical menopause and HRT use, included a measure of change, and used an age variable adjusted for MTS. Kaufert and colleagues were the only researchers to present multi-point longitudinal patterns of depression scores, and they used the shortest time interval between data collections as well. In addition, they included epidemiologically important categories of surgical menopause and HRT users. They also measured change. Dennerstein and colleagues were one of the few research teams to emphasize wellness and well-being not just illness and disease. To date they are also the only researchers of longitudinal depressive symptoms in midlife to report assessment of hormonal levels as well as MTS. Hunter used a psychological theory of depression to investigate predictors and covariates in depth. Holte was the single researcher to use multi-point data collection and an analytic technique that accommodated inter-individual correlation of repeated measures. The Gothenberg study was able to determine psychiatric diagnoses by expert interview, included all mental health disorders and associated levels of disability, and to determine incidence rates during midlife. The latter were particularly important in terms of longitudinal patterns of depressive symptoms during midlife.

No single study of depression during the transition to menopause has been without strengths and weaknesses. The study that remains to be done would combine strengths from the examples reviewed here while avoiding their weaknesses. This study's characteristics would be: 1) adequately-sized population-based sampling, 2) measurement of covariates and predictors related to depression, such as history of depression, use of therapies for depressive symptoms, life events and stressors, including history of violence 3) use of evidence-based definitions of MTS applied to prospectively

collected data, 4) measurement at multiple time points spaced both closely enough and over a long enough time period to capture change in both depressive symptoms and MTS, and 5) use of appropriate analytic strategies for repeated measures.

### Significance

If both course and causes, or at least predictors, of depression were better understood, midlife women who were depressed might be provided with more appropriate therapies and interventions. In particular, if aspects of the menopausal transition and/or menopause are causal or contribute to depressive symptoms in midlife, primary prevention, targeted screening and early diagnosis, and especially efforts to shorten the course and prevent recurrences of depression might be more effectively designed. For example, HRT or ERT have become popular approaches to treatment of depressive symptoms in midlife women (Burt et al., 1998; Schmidt et al., 2000). So far as psychosocial aspects of menopause, various “beauty” products are vigorously marketed to prevent the appearance of aging in women (Stoppard, 2000, pp. 174, 180, 210), a worry that might lead to depression. Increased social activities or employment have been proposed for empty nest syndrome (Stoppard, 2000, Chap. 9; Wilbur & Dan, 1989).

However, if depressive symptoms in midlife are not related to the menopausal transition and menopause, a different focus might be adopted, a focus on depressive symptoms rather than on menopause. If, for example, depressive symptoms are related to life events and stressors, then understanding the relationships between them would be helpful in designing interventions. If depressive symptoms in midlife women are related

to previous depressive symptoms, then an understanding of the longitudinal pattern of depressive symptoms would be necessary to predict and hopefully intervene effectively during this time of life for women.

The SMWHS can fill many gaps and avoid many pitfalls in the research literature on longitudinal depressive symptoms during the menopausal transition. SMWHS measured depressive symptoms both frequently enough (every year) and for a long enough time period (up to 10 years) for changes in both MTS and depressive symptoms to be determined. Measurements of possible causal factors were also available and most data had been recorded prospectively. The sample was from the Seattle community, not clinical practices or diagnosed patients, and was fairly large, over 200 women providing the multi-time point data. MTS had been determined prospectively and precisely for the participants. Women who had hysterectomy without bilateral oophorectomy or who used HRT were included. Analytic methods could be applied that capitalized on the multi-time point data set while not violating statistical assumptions. These factors provided an excellent opportunity to explore and describe longitudinal patterns of depressive symptoms in a community-based sample of midlife women during the menopausal transition.

**Table 1**  
**Longitudinal Studies Of Depression In Midlife Women:**  
**Study Name And Date**

<b>Name of Study</b>	<b>Site of Study</b>	<b>Senior Researcher</b>	<b>Co-authors</b>	<b>Date published</b>
Seattle Midlife Women's Health Study	Seattle, WA	N. F. Woods	Mitchell	1996
University of Pittsburgh Healthy Women Study	Allegheny county, PA (includes Pittsburgh)	K. Matthews	Bromberger	1990
Massachusetts Women's Health Study	Massachusetts state	S. McKinlay	Avis, Brambilla, Vass	1994
National Health Examination Follow-up Study	U.S.	P. Costa, Jr.	Busch, Zonderman	1994
Manitoba Project on Women and Their Health in the Middle Years	Manitoba province, Canada	P. Kaufert	Gilbert & Tate	1992
South-East England Study	Southeast England, (London and vicinity)	M. Hunter		1990, 1992
Melbourne Women's Midlife Health Project	Melbourne, Australia	L. Dennerstein	Dudley & Burger	1997
Norwegian Menopause Project	Oslo, Norway	A. Holte		1992
Longitudinal Study of Women in Gothenburg	Gothenburg, Sweden	T. Hällström	Samuelsson	1985

**Table 2**  
**Menopausal Status In Nine Longitudinal Studies of**  
**Depressive Symptoms In Midlife Women**

SITE OF STUDY, Senior Investigator	MENOPAUSAL STATUS LABELS	NUMBER OF SUB- CATE- GORIES	DEFINITIONS
Seattle, Woods (1996)	Not given	Not given	<u>Asked changes in interval and regularity within last 6 months</u>
Allegheny County, Matthews (1996)	Premenopausal  Peri- or postmenopausal  HRT	3	Menses in last 3 months  Ceased menses for last 3-12 months  Hormone Replacement Therapy (incl. Post-hysterectomy)
Massachusetts, McKinlay (1994)	Premenopausal  Perimenopausal  Postmenopause (Natural menopause)	3	<u>Asked changes within last 9 months</u> Regular menses last 3 months  Periods of amenorrhea and/or changes in regularity or flow past 12 months  No menses last 12 months
U.S., Costa, Jr. (1994)	Premenopausal  Perimenopausal  Natural menopause  Surgical menopause	4	<u>Based on change from 10 years earlier when regular menses</u> Pregnant or regular menses  Irregular menses due to the change of life  No longer menstruating naturally ( $\geq 12$ months)  No longer menstruating r/t surgery
Manitoba, Kaufert (1992)	Premenopause  Perimenopause  Postmenopause  Hysterectomy	4	<u>Asked changes during last 6 months</u> Regular menstruation  Other than pre, post or hysterectomy  No periods x 2 consecutive assessments (= 12 months)  With or without oophorectomy

Table 2 continued

SITE OF STUDY, Senior Investigator	MENOPAUSAL STATUS LABELS	NUMBER OF CATE- GORIES	DEFINITIONS
Melbourne, Dennerstein (1997)	Premenopausal	5	No change in frequency or flow within last 3 months
	Early perimenopausal		Some change in frequency and/or flow within last 3 months
	Late perimenopausal		3-11 months amenorrhea
	Postmenopausal 1		1-2 years amenorrhea
	Postmenopausal 2		>2 years amenorrhea
South-East England, Hunter (1990)	Premenopausal	3  (NB: peri and post combined for analysis)	<u>Change from premenopausal at entry</u> Regular menses last 12 months
	Perimenopausal		Irregular menses (cycle >6 weeks)
	Postmenopausal		No menses x 12 or more months
Oslo, Holte (1992)	Premenopause	2	≤6 months amenorrhea
	Postmenopause		>6 months amenorrhea
Gothenburg, Samuelsson (1985)	Preclimacteric	5	38 years old
	Premenopausal		regular menstruation (no amenorrhea x ≥ 2 months)
	Perimenopausal		beginning irregular menses (=skipping menses)
	Early postmenopausal		amenorrhea 12-35 months
	Late Postmenopausal		≥3 years amenorrhea

Table 3  
 Longitudinal Studies Of Depression During Menopausal Transition: Methods

Site of Study (P.I.)	Sample size	Sampling method	Calendar years of data collection	Number of time points of measurement	Depression Measurement tool	Other independent variables measured	Age at entry to study	Analysis methods
Seattle (Woods)	347	Multistage, community-based telephone numbers	1990-92; 12 months later	2 1 year apart	CES-D*	7 from factor analysis: Social support Feminine socialization Health Vasomotor Stress Family Resources Socialization midlife	35-55	Discriminant analysis 4-group outcome variable: Absent depressed mood Consistent depressed mood Emerging depressed mood Resolving depressed mood
Allegheny County, PA including Pittsburgh (Matthews)	460	Random drivers licenses, selected zip codes; healthy at entry only	1983-85; 1986-88	2 3 years apart	BDI**	Hx stressor last 6 mo Current stressor Pessimism Trait anxiety Depression at T1 Education at T1 HRT*** at T2	42-50	First order correlations  Stepwise linear regression
Massachusetts (McKinlay)	2352 (2347 at end of study)	Random, state census lists	1982-84; 27 months later, through 1987	2 27 months (2.25 years) apart	CES-D	Hot flashes+night sweats+menstrual probs HRT at T2 Depression at T1	46-56	Logistic regression Likelihood ratio test for differences among 5 MTS

\*Center for Epidemiologic Studies-Depression Scale

\*\*Beck Depression Inventory

\*\*\*Hormone Replacement Therapy

Table 3 continued

Site of Study (P.I.)	Sample size	Sampling method	Calendar years of data collection	Number of time points of measurement	Depression measurement tool	Other independent variables measured	Age at entry to study	Analysis methods
U.S. (Costa, Jr.)	394 (sum for CESD calculated, not given)	Stratified probability of adult, non-institutionalized civilian population of U.S.	1971-75; 1982-84	2 10 years apart	CES-D	Age within menopausal transition stage Years since menopause	40-58	Repeated measures ANOVA 2-way MANOVA
Manitoba, Canada (Kaufert)	469 145 post hysterectomy at entry 324 no hysterectomy at entry	Non-random mail survey selected from general population of women in age group in province	1981 1982 1983 (est. from publication dates, incl. pilot study)	5 (pooled into 2) 6 months apart	CES-D	Present health Sum of major health problems last 6 mo Family events last 6 mo: death, children leave Current stress: probs with fam, relationships, other	45-55	Chi-square of paired scores from one time point to the next time point (5 time points pooled into multiple pairs) Logistic regression with backward stepwise entry of independent variables

Table 3 continued

Site of Study (P.I.)	Sample size	Sampling method	Calendar years of data collection	Number of time points of measurement	Depression measurement tool	Other independent variables measured	Age at entry to study	Analysis methods
Southeast England (Hunter)	36	Volunteer clinical	1983; 1986	2 3 years apart	7-item scale derived from factor analysis of 36 symptom set	Age within menopausal transition stage SES**** (2 groups) Marital status Employment status Depression at T1 Menopause beliefs Expectations Stress (yes/no) Hypochoondriasis Regular exercise	47-55	Paired t-tests and McNemar tests  Multiple stepwise regression
Melbourne, Australia (Dennerstein)	438 (405 at end of year 4)	Random, community-based, Australian-born	1991; 1992; 1993; 1994	4 (pooled) 1 year apart	Affectometer 2	Age adjusted for menopausal trans stage Free estradiol Estradiol index Free FSH Free androgen index Hot flashes (bothersome, last 2 wks)	45-55	Generalized linear model, pooled data set assuming normal errors, stepwise entry of variables

\*\*\*\*Socioeconomic status

Table 3 continued

Site of Study (P.I.)	Sample size	Sampling method	Calendar years of data collection	Number of time points of measurement	Depression Measurement tool	Other independent variables measured	Age at entry to study	Analysis methods
Oslo, Norway (Holte)	59	Random, Community Register of Central Bureau of Statistics	1981; 1982; 1983; 1984; 1985; 1986	5⇒2 dichotomized to pre and post menopause 1 year apart	4-item yes/no scale from GHQ <sup>†</sup> + 3 ques for each item: 1. Current yes/no 2. Freq in last 12 mo (12-pt) 3. Amt distress (4-pt)	Age Age within Menopausal transition stage Vasomotor sx	47-56	T-test of summary measures pre vs post menopause
Gothenberg, Sweden (Hallstrom)	677	Systematic, Taxation Office Register	1968-69; 1974-75	2 6 years apart	DSM-III <sup>††</sup> criteria, semi-structured psychiatric interview	Age at entry Marital status No. of children Social class Employment (0,PT,FT) Weighted life events (complex measure)	38-54	Nonparametric comparisons (Pitman's permutation test, Fisher's permutation test, or Fisher's exact test) of percentages or rates, with Mantel method of adjustment for confounding variables

<sup>†</sup>General Health Questionnaire

<sup>††</sup>Diagnostic and Statistical Manual, 3<sup>rd</sup> ed.

### **CHAPTER 3: METHODS—PARENT STUDY and GENERAL METHODS**

This chapter reviewed research design and methods of the parent study and design and methods applicable to both analyses conducted in the current study. The Seattle Midlife Women's Health Study (SMWHS), Nancy Fugate Woods Principal Investigator, and Ellen Sullivan Mitchell Project Director and Co-Investigator, has been referred to here as the parent study. The parent study is based in the University of Washington, School of Nursing, Department of Family Child Nursing. Since its beginning in autumn 1989, this researcher has worked nearly continuously as Research Assistant, Research Coordinator, Research Associate, and eventually Research Scientist and Co-Investigator. Starting as an original Research Assistant in September 1989, the author personally interviewed 173 women, then expanded tasks on the parent study to include data management and processing for five years until Oct 31, 1994. Continued data management and processing intermittently over the next three years segued into a return in November, 1997 to full-time responsibility for direct caseload management, again including personal contact with the participants to collect data, as well as data management and analysis.

From the administrative point of view the parent study may be divided into three time periods relating to the grant support received. The first grant, titled "Midlife Women: Health and Health Seeking Behaviors" and funded by the National Institute of Nursing Research<sup>1</sup> (NINR) of the National Institutes of Health (NIH), supported the five years from September 1989 through October 1994 in the Center for Women's Health

---

<sup>1</sup> P30-NR04001

Research of the University of Washington School of Nursing. Smaller Intramural Research Support Grants (IRSG) were obtained for 1994-5 and 1995-6, titled "Bleeding Patterns, Symptom Experiences and Perimenopausal Phase in Midlife Women" and "A Dynamic Analysis of Symptom Patterns Across the Menopausal Transition," respectively. The third and current grant, funded again by the NINR<sup>2</sup>, started in September 1996 and continued through May 2001. It is titled "Menopause Transition: Biobehavioral Models of Symptoms." Each grant, and therefore each time period, had somewhat different aims and as a result some methods differed between periods. However, the parent study has always included as an aim the observational, longitudinal description of the natural history of transition to menopause and the first years after menopause.

#### Population

The study population consisted of English-speaking Seattle-residing midlife women who had not yet completed the transition to menopause at entry to study and for whom the transition could be studied.

#### Sampling Frame

The sampling frame was comprised of all residential phone numbers in selected census tracts chosen for mixed ethnicity and income, based on U.S. Census Bureau information, and that were in a purchased reverse phone directory for Seattle, WA in late 1989.

---

<sup>2</sup> RO1-NR04141-01A1

## **Sampling Procedure**

### **Contact Procedures**

Telephone calls to every residential phone number in the census tract were attempted up to nine times, varying the time of day and day of week if needed. A telephone script was followed and if the respondent was eligible and had verbally consented to an initial personal interview one was arranged, usually at the respondent's home, although occasionally at the University of Washington.

### **Eligibility/Exclusion Criteria**

Midlife women (35-55 years old) were eligible if they had had at least one menstrual period within the preceding 12 months, still had a uterus and at least one ovary, were not pregnant or lactating at time of entry, could read and speak English, and had at least some high school education. Throughout the duration of all study procedures, beginning in 1990 women became ineligible after bilateral oophorectomy. Women were also considered ineligible if they had not provided information the previous year unless special arrangements were made for delayed participation.

At the beginning of the second NINR grant in November 1996, women were still eligible if they had provided at least two years of data during the first NINR grant. Some women who had previously dropped out of the longitudinal portion of the study subsequently re-entered. Women were also required to be no more than five years postmenopause at entry to the second NINR grant. During the second NINR grant, women became ineligible if both ovaries were removed, or at five years postmenopause, or at age 60 years if also using HRT for more than two consecutive years. In 1999,

women who had discontinued completion of questionnaires were provided with the option of completing an abbreviated version, and again some women who had previously dropped out of the longitudinal portion of the study subsequently re-entered.

### **Sampling Results Of Parent Study**

The total number of households screened was 11,222. From January 1990 to January 1992, 508 women were interviewed. At entry to the SMWHS racial and ethnic distributions of the sample reflected those of Seattle (Woods & Mitchell, 1997), 77% Euro-American, 11% African-American, and 7% Asian-American. Women in the sample at entry had a median education of 15 years, with a median household income of \$38,000 in 1991 dollars; at the time 68% were partnered. At the beginning of the second year of participation 367 women were still providing adequate information to remain enrolled. At entry to the second NINR grant period 277 women were still eligible after contact and 221 provided initial information. Racial and ethnic distributions at entry to the second NINR grant were 85% Euro-American, 6% African-American, and 8% Asian-American, with a mean education of 16 years, median household income of \$40,000 in 1997 dollars; at the time 71% were partnered.

### **Tools And Measures**

#### **Center For Epidemiologic Studies-Depression Scale**

The Center for Epidemiologic Studies-Depression scale (CES-D) is a 20-item self-report measure of depressive symptoms derived from clinical criteria for MDD in the categories of positive affect, negative affect, somatic, and interpersonal symptoms (Radloff, 1977). Participants are asked to rate how often they had a specific feeling

during the past week on a 4-point scale from 0 (rarely or none of the time (less than 1 day)) to 3 (most or all of the time (5-7 days)). The four items of the positive affect category are worded positively and are reverse coded (see Appendix A). The total CES-D score is a sum of the ratings of the 20 items, ranging from 0 to 60. The CES-D was developed in the 1970s for use with community samples and the first reports on its development, reliability, and validity were published by Radloff in 1977. In her various samples test-retest reliability correlations were .40 or above, and internal consistency reliability as measured by Cronbach's  $\alpha$  was .80 or above.

Since the 1970s the CES-D has been widely used in health research as a measure of depression, thereby allowing comparisons across studies (Furukawa, Hirai, Kitamura, & Takahashi, 1997). However, the CES-D is not a diagnostic tool (Boyd, Weissman, Thompson & Myers, 1982; Boyle, 1985; Breslau, 1985; Craig & van Natta, 1976; Devins & Orme, 1985; Faulstich, Carey, Ruggiero, Enyart & Gresham, 1986; Myers & Weissman, 1980; Roberts, Lewinsohn, & Seeley, 1991; Roberts, Rhoades, & Vernon, 1990; Roberts & Vernon, 1983; Roberts, Vernon, & Rhoades, 1989). It has been found to measure distress or demoralization rather than MDD per se, and in particular cannot distinguish depression from anxiety (Breslau, 1985; Devins & Orme, 1985; Roberts et al., 1990). While in general no ethnic or language (when translated) effects for the CES-D have been found (Cho, Moscicki, Narrow, Rae, Locke & Regier, 1993; Furukawa et al., 1997; Roberts et al., 1989; Roberts et al., 1990; Somervell, Beals, Kinzie, Boehnlein, Leung & Manson, 1993), some studies have found age differences (Beekman, Van Limbeek, Braam, De Vries, & Van Tilburg, 1997; Clark, Aneshensel, Frerichs &

Morgan, 1981; Cole, Kawachi, Maller, & Berkman, 2000; Gerety, Williams, Jr., Mulrow, Cornell, Kadri, Rosenberg, et al., 1994; Guarnaccia, Angel & Worobey, 1989; Liang, Van Tran, Krause, & Markides, 1989), although several studies have found the CES-D as useful in adolescents as in adults (Garrison, Addy, Jackson, McKeown & Waller, 1991; Gjerde et al., 1988; Roberts et al., 1991). Since the current study is limited to midlife women, age effects are not a concern. There has been some question about confounding of CES-D scores from physical symptoms related to medical illnesses (Devins & Orme, 1985); however, at least one study using the CES-D has shown no such confounding (Foelker, Jr. & Shewchuk, 1992). Gender differences in total scores and gender differences in CES-D items have been found (Cho et al., 1993; Cole et al., 2000; Guarnaccia et al., 1989; Stommel, Given, Given, Kalaian, Schulz, and McCorkle, 1993), although comparisons to men's CES-D scores is not an issue for the current study. Concerns about factor structure (Clark et al., 1981; Radloff, 1977; Roberts et al., 1989; Sheehan, Fifiield, Reisine & Tennen, 1995; Zich, Attkisson, & Greenfield, 1990) are not an issue when using the total CES-D score.

The CES-D was originally designed to be used as a dichotomous variable with a cut-off score of  $\geq 16$  as indicative of major depression. Since then, various studies have investigated the CES-D cut-off score of  $\geq 16$  and found other cut-off scores indicative of major depression in different samples. In general, people who are experiencing a major depressive episode score  $\geq 16$ ; nevertheless, false positives of people not experiencing an episode but scoring  $\geq 16$  are typical (Furukawa et al., 1997; Garrison et al., 1991; Lewinsohn & Teri, 1982; Roberts et al., 1991; Zich et al., 1990) and there are false

negative rates as well (Devins & Orme, 1985). A higher cut-off, at  $\geq 24$  for girls, was found in one study of adolescents by Roberts et al. (1991). Finally, total CES-D scores and individual item ratings, especially in community samples, are not normally distributed (Avis et al., 1994; Clark et al., 1981; Devins & Orme, 1985; Sheehan et al., 1995; Stommel et al., 1993). The non-normal distribution should be taken into account when conducting analyses.

While the CES-D score was originally designed as a dichotomous variable consistent with the threshold concept of depressive disorder (Angst & Merikangas, 1997), it was decided for the current study that a simple dichotomy would lose information and that it would be more useful for longitudinal pattern analysis if the CES-D score were considered as an interval variable. In fact, for regression of simple lines and concomitant determination of slopes of the regression, the total CES-D score was considered a continuous variable.

The rationale for this decision was based on the fact that use of the CES-D for repeated measures to identify longitudinal change within the individual is not well served by a dichotomous variable (Bryk & Raudenbush, 1992, p. 130-1). Originally, a cut-off score of  $\geq 16$  was developed to discriminate non-depressed from depressed people cross-sectionally (Radloff, 1977). Longitudinally, however, the amount of decrease (or increase) in symptoms as reflected in the total score has more value than simply the change of a total score from above or below a cut-point. For example, if a woman's score changed from 56 to 28 her symptoms would have been halved. If a woman's scores changed from 7 to 15 her symptoms would have more than doubled. However,

with a dichotomous CES-D score, neither of these women would register as having changed score. On the other hand a woman whose score changed from 15 to 16 would be counted as having changed from non-depressed to depressed. Most importantly for the current study, women with an otherwise stable level around 15-17 would disproportionately contribute to any conclusions about changes in depression even though they might not be particularly unstable (see Chapter 5). Weissman, Sholomskas, Pottenger, Prusoff and Locke (1977) found that decreases in CES-D scores were clinically significant regardless of the relationship to a cut-off value. The stability of the CES-D as a repeated measure was investigated by Sheehan et al. (1995) who did find stability of the measurement structure across time, appropriate for studies of change. In order to assist the reader in viewing graphs, reference lines at the original, lowest cut-off of 16 and at the higher cut-off of 24 have been placed in line graphs.

#### Menopausal Transition Stage (MTS)

Mitchell et al. (2000) have provided an evidence-based system of MTS (see Chapter 2). In this system, theoretical considerations separate phase from stage, and consider stage independent of use of steroidal hormones or surgery. Operationally, however, all phases, stages, HRT use, and hysterectomy were coded within the MTS rubric as mutually exclusive. The following coding of the MTS variable was employed:

1. Pretransition (P) referred to the calendar year during which a woman was older than 35 years and reported no changes whatsoever in her menstrual cycle or flow from her earlier years.

2. **Early transition (E)** referred to the calendar year during which a woman reported changes in menstrual flow or regular change in cycle length after age 35 years.
3. **Middle transition (M)** referred to the calendar year after age 35 years in which a woman's prospectively recorded menstrual bleeding demonstrated a new irregularity, that is, at least one menstrual cycle that was either longer and/or shorter by six or more days compared to a contiguous cycle.
4. **Late transition (L)** referred to the calendar year after age 35 years in which a woman's prospectively recorded menstrual bleeding demonstrated skipping, that is, at least one menstrual cycle was twice the years' modal cycle length.
5. **Postmenopause (PM)** referred to the calendar year during which no menstrual bleeding or spotting was prospectively recorded with no other explanation for the amenorrhea.
6. **Hormone use (HRT or H)** referred to the calendar year during which a woman used hormone replacement therapy for at least four of the 12 months, whether she was known to be postmenopausal or not. Because of the effects of HRT on menstrual bleeding and the ovary, this category was operationalized as mutually exclusive with the others. Nevertheless, theoretically women using HRT were also either P, E, M, L, or PM. In practice, women were most likely to be L or PM.
7. **Oral contraceptive use (OC)** referred to the calendar year during which a woman used oral contraceptives, including progestin-only formulations, for at least four of the 12 months, again regardless of the theoretical underlying MTS.

8. "Other" included all women who could not be classified in one of the above categories, such as women post-hysterectomy who retained at least one ovary. This category included Unknown (U), which referred to the calendar year during which a woman's transition stage could not be determined. Often this was due to incomplete recording on her part. Many women were U at entry to the study because they had not recorded and could not remember menstrual cycle information well enough in order to determine stage before the study.

#### Antidepressant And Tranquilizer Medications

Tricyclic antidepressants (TCA), monoamine oxidase inhibitors (MAOI), selective serotonin reuptake inhibitors (SSRI), selective norepinephrine reuptake inhibitors (SNRI), and lithium are considered antidepressants, regardless of their intended purpose. Some antidepressants have been marketed under different names recently for smoking cessation (bupropion no longer as Wellbutrin but now as Zyban) and for cyclic short-term use for premenstrual dysphoric disorder (fluoxetine no longer as Prozac but now as Serafem). For some women antidepressants have been used for chronic pain treatment and other diagnoses such as eating disorders.

Use of antidepressants, major tranquilizers and sedatives was recorded in diaries 1, 2, and 3. Diaries 1, 2, and 3 corresponded to the first three occasions of CES-D measurement. Because there was no diary at exit from the first NINR grant, there was no corresponding information available about medication use around the time of the fourth CES-D measurement. Use of antidepressants, major tranquilizers and sedatives continued to be recorded in diaries from approximately January 1995, January 1996,

January 1997, corresponding to later CES-D measurement. Starting in 1998 all medications taken regularly within the preceding month were recorded in the same questionnaire containing the CES-D scale. At every measurement occasion, all women taking major tranquilizers and sedatives were also taking antidepressants; therefore, use of all or any of these medications was coded and was referred to as antidepressant use alone.

However, medication data in the computerized data file from the first three diaries were unfortunately not necessarily available for the first day of recording in the diary, corresponding to time of CES-D measurement. The computerized diary data set was limited to the first time in the daily recordings that a woman was premenstrual, five days before the onset of menses. It was possible and even probable, that several days or even weeks of recording in the diary were not available. While a change in the use of medications could have occurred during the time between completion of the CES-D and the corresponding medication recorded in the diary, this change was judged not to be likely or common for antidepressants or major tranquilizers and sedatives. Use of antidepressants in particular is usually not expected to be intermittent or haphazard. That is, once a woman has been prescribed an antidepressant she is likely to take it daily for several months at least. It was therefore assumed that use of antidepressants on any day in the diary would likely reflect continuous use on nearby days not data entered.

St. John's wort, or *Hypericum*, was not coded as an antidepressant, although there were noted to be quite a few women using this herb for mood improvement in late 1998 after considerable attention in the media. Also, no distinguishing codes were available to

separate lithium, for bipolar depressives, from other antidepressant medications. TCAs, SSRIs, SNRIs, and MAOIs were coded as antidepressants regardless of the purpose for which they were prescribed, such as smoking cessation.

### Study Protocol

At the personal interview at entry to the study (1990-1992) written consent was obtained, study procedures were explained, and administration of the CES-D was conducted as a card sort. After the initial interview the CES-D was administered as part of a mailed questionnaire approximately annually.

At entry to the study the total score for the CES-D was given the variable name CESD, corresponding to diary 1. Annual questionnaires containing the CES-D were then administered at approximately 12 (corresponding to diary 2), 24 (corresponding to diary 3), and 36 months after entry to the study and given variable names CESD12, CESD24, and CESD36, respectively. Because entry to the study required 25 months for all 508 women, and because not all questionnaires were completed and returned on schedule, women were completing the CESD36 throughout 1993 and 1994. Early-entry women, who had completed CESD36 in early 1993, were therefore not administered a CESD in 1994. Even though they might have completed every CES-D administered, they nevertheless skipped a calendar year, 1994.

On the other hand, in 1995 the entire sample was converted to a January questionnaire. Women whose CESD36 had been completed within the previous six months (second half of 1994) did not receive CESD95. In data files, therefore, CESD95 appears as missing for these women even though no questionnaires were omitted. From

1996 on, repeated annual total CES-D scores were distinguished by adding the year of administration to the variable name, e.g., CESD96, CESD97, CESD98, CESD99, and CESD00.

#### Sampling Frame Of Current Study

The sampling frame of the current study consisted of all CES-D scores provided by all participants in the SMWHS. Descriptive statistics for each time of administration of the CES-D in the parent study have been provided in Table 4. Decrease in sample size during the years between NINR grants and trend to decreasing total CES-D scores over time can be readily seen. Increase in sample size in 1997 was due to re-entry of women that had dropped during the interval between NINR grants, and increase in 1999 was due to re-entry of women that had discontinued completing full questionnaires and were provided the option of completing an abbreviated version that contained the CES-D.

#### Measurement Procedures Of Current Study

For the current study, total CES-D scores were re-named so that a T preceded the variable name, and in the case of the first score a suffix was added (e.g., TCESDt0 for time zero, TCESD12, TCESD24). Every missing CES-D item was reviewed on the original questionnaire documents. A conservative approach was taken to imputing a total CES-D score if there were missing items. Only one item was allowed to be missing, and then only if it could be reasonably attributed to the formatting of the typing on the page (e.g., two consecutive items single spaced, the anchor/introduction to the scale typed too close to the first item). Total CES-D scores for three women, all at time 12 months (CESD12), were therefore made entirely missing. In addition, three scores for another

woman were imputed, CESD12, CESD97 and CESD99. There were nine (9) other women who had a single total score adjusted by this procedure, over a random distribution of years, excluding CESD24.

#### **Ethical Considerations**

The procedures of the SMWHS and consent forms describing purpose, benefits, risks, procedures, confidentiality, and option to withdraw without penalty have been approved by the Human Subjects Review Committee of the University of Washington #96-1669-C/E 07 (see Appendix B). After description of study procedures at entry to the parent study, again at exit from the first NINR grant and entry to the IRSG phase, and lastly at entry to the second NINR grant, participants provided written consent to participate in each funding phase of the SMWHS.

**Table 4**  
**Descriptive Statistics**  
**Yearly Total CES-D Scores From the Seattle Midlife Women's Health Study**

	TCESD10	TCESD12	TCESD24	TCESD36	TCESD95	TCESD96	TCESD97	TCESD98	TCESD99	TCESD00
N	508	344	307	249	141*	189	220	201	211	193
Mean	11.73	11.76	11.56	11.33	11.23	11.06	10.95	9.95	10.70	9.48
Std. Deviation	8.62	8.97	8.62	8.75	8.53	9.35	9.47	8.54	9.23	8.81
Percent at or above score of 16	24	28	24.4	21	25	26	22	17	23	17
Percent at or above score of 24	9	10.5	11	8	7	11	9.5	4.5	8.5	6
Cronbach's alpha	.878	.903	.896	.904	.892	.906	.914	.894	.907	.903

\*CESD95 administered to only those women who had not completed CESD36 within six months of January 1995

## **CHAPTER 4: CLUSTER ANALYSES**

This chapter reported research aims, specific design, methods, results, and focused discussion for a series of cluster analyses and predictors of clusters results, including a review of the analytic methods in case these were not familiar to the reader. The first section reported k-cluster analysis, the second section reported hierarchical agglomerative cluster analysis, and the final section contained analysis of predictors of cluster results. General information about the parent study and general methods for the current study have been provided in Chapter 3.

### **Aims**

1. The primary purpose of this study was to describe and characterize longitudinal patterns of depressive symptoms in midlife women before, during, and after the transition to menopause.
2. The secondary purpose of this study was to explore relationships between longitudinal patterns of depressive symptoms and selected predictors of these patterns or particular changes within patterns. Specific predictors were a) MTS and b) history of harassment, abuse, or assault. Antidepressant drug use was considered a potential confounder.

### **Research Design**

The current study was conducted with data from the SMWHS. It was exploratory and descriptive in design. Description was focused on the longitudinal pattern of repeated measures over time. The time interval between repeated measures was approximately one year and the number of time points was six or more, up to 10.

## Methods

In order to achieve the exploratory and descriptive aims of the study, non-inferential methods were required. In addition, methods had to be designed for or adaptable to description of longitudinal patterns of repeated measures over time. One of the difficulties with analyzing longitudinal data sets containing three or more time points is the complexity and/or obscurity of the statistical approaches available (Collins & Horn, 1991, see also Chapter 2). While not usually conceptualized as a statistical method for longitudinal pattern analysis, cluster analysis (Aldenderfer & Blashfield, 1984) can be used for this purpose. Cluster analysis is readily available, is relatively easy to understand and use, and produced meaningful results regarding longitudinal patterns.

Cluster analysis does not assume any relationship between the variables submitted for clustering. Common examples of clustering (Aldenderfer & Blashfield, 1984; Cluster analysis, 1997, Chap. 15; Hierarchical cluster analysis, 1997, Chap. 25-26) do not analyze variables that are themselves correlated. That is, variables submitted for clustering are usually independent. In the current study, however, the variables were actually repeated measures; each variable was the annual CES-D score, being repeated purposefully to describe patterns of depressive symptoms over time. A longitudinal set of repeated measures demonstrates at least intra-individual correlation and likely even autoregression. In order to improve the clustering of patterns over time, more than just individual annual CES-D scores, variables reflecting the correlated and autoregressive nature of patterns of repeated measures were also submitted for clustering. These variables have been described in the appropriate methods sections that follow.

While cluster analysis is usually conducted as a single analysis, what follows are selected results from a set of cluster analyses. The set was conducted because a single cluster analysis might have been inadequate to fully explore, validate, and determine reliability for the unusual data set of longitudinal patterns of repeated CES-D scores over time. Therefore, as many types and approaches to cluster analysis as were reasonably possible were performed.

### Review of Analysis Methods

Modern cluster analysis has been available for several decades and used in both the biological and social sciences. There are several approaches, yet all have the same purpose: to group individuals or cases that are similar in relation to a set of variables (Aldenderfer & Blashfield, 1984, Chap. 1, p. 33).

Cluster analysis has several parameters that must be specified, depending on theoretical and practical bases. The first parameter is the type of cluster analysis. Choice of type depends primarily upon the total sample size, as well as the strengths and weaknesses of the different types of cluster analysis. Iterative clustering or k-cluster analysis, also called Quick Cluster by SPSS, is generally used for sample sizes larger than 200 (Aldenderfer & Blashfield, 1984; Cluster Analysis, 1997). It also allows for missing data of an individual variable, retaining the case for clustering, if the subcommand is specified as pairwise. However, “unfortunately, there is really no objective way to determine if a solution obtained from an iterative partitioning method is globally optimal” (Aldenderfer & Blashfield, 1984, p. 48). On the other hand, hierarchical agglomerative cluster analysis is appropriate for smaller sample sizes, although it requires there be no

missing data. Both types of cluster analysis were used in the following set of analyses in order to maximize the benefits of cluster analysis in general and also balance out the weaknesses of the different methods.

The second parameter that needs to be determined is the number of cluster solutions specified for the analysis. The number of clusters should be related to both theory concerning the topic of interest and the number of variables submitted for clustering. The third parameter to be decided is the selection and number of variables to be submitted for clustering. Such decisions are grounded in theory concerning the topic of interest.

Lastly, when hierarchical agglomerative cluster analysis is used, both the method and the measure must be specified from a range of choices. Various methods of grouping similar cases into clusters, referred to as linkage, are available (Aldenderfer & Blashfield, 1984). Ward's method joins cases that "result in the minimum increase in the error sum of squares" (p. 43). Ward's "has been shown to outperform most other clustering methods" (p. 61) and "tends to find (or create) clusters of relatively equal sizes and shapes as hyperspheres" (p. 43).

Between-groups Average Linkage is a method that "computes an average of the similarity of a case under consideration with all cases in the existing cluster and, subsequently, joins the case to that cluster if a given level of similarity is achieved using this average value" (Aldenderfer & Blashfield, 1984, pp. 40-41). Complete Linkage is a method that requires a case be "within a certain level of similarity to all members of that

cluster” in order to be added to that cluster. Similar to Ward’s method, it tends to find “relatively compact, hyperspherical clusters composed of highly similar cases” (p. 40).

There are various formulas available to measure the similarity between two cases (Aldenderfer & Blashfield, 1984, p. 16-33). Three measures were considered useful, Squared Euclidean, City Block or Manhattan, and Chebychev. Squared Euclidean is the sum of the squared differences between two cases of the values of the clustering variables. City Block is the sum of the absolute differences between two cases of the values of the clustering variables. Chebychev is the maximum absolute difference of the values of the clustering variables (Cluster, 1997). What follows for each type of cluster approach is:

1) **Methods.** A description of the various parameters that were used followed the same order as in the review of the method above: First the type of cluster analysis is described, then the number of cluster solutions that were specified, followed by the selection of variables, and finally for hierarchical clustering, the method and the measure.

2) **Results.** Presentation has been limited to the one cluster result of each type that was deemed the best. Keeping in mind the study aim of describing longitudinal patterns, as well as the clustering purpose of grouping individuals or cases that are similar in terms of a set of variables, the result selected as best demonstrated the greatest number of distinct patterns with the fewest number of clusters. Consideration was given to the variability or heterogeneity within a cluster as well as clear distinction between clusters, and to the sense of longitudinal pattern demonstrated by each cluster.

3) **Summary.** A capsule of results is given, including those analyses conducted but not presented.

4) **Discussion.** Interpretation of cluster results, limitations, and implications for further analysis are discussed.

### Sampling Procedure

The sampling frame consisted of all women participating in the SMWHS, or the parent study. Women with at least 3, 4, 5, and 6 annual CES-D scores through 1999 were considered for clustering. Requiring at least three total CES-D scores was originally judged as a minimum requirement for longitudinal pattern determination. The advantage of requiring the lower number of repeated measures was to increase the sample size and not lose possible patterns.

However, there was some evidence from other statistical approaches such as hierarchical linear modeling that at least five repeated scores might be necessary for longitudinal data analysis, according to a statistical consultant (Robert Abbott, personal communication, Feb. 2000). In addition, on advice of the statistical consultant, the minimum number of longitudinal data points should be no fewer than the harmonic mean of the number of annual CES-D scores per woman. The harmonic mean from the SMWHS sample of women with at least three CES-D scores was 5.8. Phenomenological analysis (see Chapter 5) also required at least six longitudinal data points in order to determine patterns. The consistency of these different sources in requiring at least five or six data points supported a decision to use women with at least five or six repeated total CES-D scores. Requiring at least six scores provided a sample of adequate and

appropriate size for iterative clustering and enough repeated measures to determine longitudinal patterns. In the end, however, the decision to use women with at least six (or five for hierarchical clustering) CES-D scores was primarily arbitrary.

### **ITERATIVE OR K-CLUSTER ANALYSIS of YEARLY CES-D SCORES PLUS SLOPE99**

#### **Analytic Procedures**

##### **Type Of Cluster Analysis**

The first approach to cluster analysis used iterative or k-clustering.

##### **Number Of Cluster Solutions**

The number of cluster solutions was arbitrarily specified as 7, 9, 11, and 13, all using the same variables and sample.

##### **Selection Of Variables**

Each annual CES-D total score was submitted as a separate variable for clustering in early cluster analyses. Scores were available through CESD99 (see Chapter 3 for description of variable names). Because of the procedural methods of timing the administration of the CES-D, some women (45% of the sample with six or more annual CES-D scores) were not administered and therefore did not complete a CESD95 (see Chapter 3). In data files, therefore, CESD95 for these women was designated as missing. Cluster analyses were performed both with and without this variable; with a few exceptions the results in cluster assignment and cluster centers differed only in minor ways. The main exception was, however, that a 1995 rise was omitted in one pattern. The variable CESD95 was therefore retained in the final list of submitted variables.

In order to improve consistency of pattern within clusters and to make the cluster results more meaningful in terms of longitudinal pattern, summary variables of each individual woman's longitudinal set of repeated CES-D scores were added to the variables submitted for clustering. These summary measures or computed variables have been referred to in all capital letters with the year-designation from the final yearly CES-D score used in the computation. Summary measures included a woman's mean of all her CES-D scores through 1999, MEAN99; the range of all her CES-D scores through 1999, RANGE99; and the slope of a simple line, SLOPE99, fit through the set of all CES-D scores of each woman regressed on time in years, that is, graphed along a weighted time axis. Weighted time was calculated by using the date written on the annual questionnaire or its postmark. The year of entry to the study was set to zero; each subsequent date of CES-D completion was calculated in years since entry to the study. Alternately to RANGE99, the summary measure of variability was submitted as the standard deviation of all her CES-D scores through 1999, SDCESD99, or the coefficient of variation of all a woman's CES-D scores through 1999, CVCESD99.

Summary measures allowed a representation of longitudinal pattern in the following ways: MEAN99 was a reflection of general level of depressive symptoms; RANGE99, SDCESD99, and CVCESD99 were representations of the overall variability or lability of depressive symptoms. The overall sense of trend over time was summarized as SLOPE99, positive if the scores were generally rising and negative if the scores were generally falling.

Addition of other variables such as MEAN99, RANGE99, SDCESD99, and CVCESD99 to the set of six or more scores did not improve the discrimination of clustering. These summary variables were understood to be redundant, decreasing the efficiency of the clustering and overwhelming true discriminating factors for clusters. On the other hand, the addition of SLOPE99 to the cluster analyses resulted in clusters with better discrimination than analyses that used only the yearly CES-D scores. Therefore, for k-clustering the variable SLOPE99 was included for submission along with each annual CES-D score.

## Results

### Sample

There were 216 women with six or more CES-D scores through 1999.

### Number Of Cluster Solutions

The best number of cluster solutions was found to be 9, submitting each woman's annual CES-D scores plus her SLOPE99 for clustering. Seven cluster solutions produced one very populous cluster that was unsatisfactorily heterogeneous. Eleven- and 13-cluster solutions tended to subdivide already less populous clusters. The rare clusters with very few women tended to change little regardless of the number of clusters specified. These few women had patterns that were so dissimilar from other women that they were not easily grouped with any of the others regardless of the specifics of clustering. In the 11- and more so in the 13-cluster solution there was also apparent clustering on an "elevation continuum" (Aldenderfer & Blashfield, 1984, p. 67) where, for example, centers in one cluster were scores of three to five and in another cluster were

six to nine. Such a distinction was not judged to be clinically meaningful. Based on the difficulties interpreting these analyses, the 9-cluster solution was chosen as best. Cluster centers for each variable of the 9-cluster solution have been presented in Table 5. Brief verbal descriptions of the longitudinal patterns were coined for each cluster.

### Clusters

A graphic representation of the yearly means of each cluster is presented in Figure 1. Two clusters, #4 and #5, contained very high scores and very high ranges, and had only two cases each. There were, therefore, seven general clusters in the 9-cluster solution that were used for identification of common patterns.

Cluster #7, the most populous cluster, was identified as Non-depressive or Non. The plurality of women in this sample (36%) were in this cluster, meaning that in the single most common pattern women were generally free of depressive symptoms over the entire six to eight or nine year period. Cluster #3, the next most populous, was titled Falling or Resolving depressive symptoms at a low level. Cluster #2, similar in number of cases to #3 and in time pattern, with negligibly lower CES-D scores and slightly flatter slopes, was labeled Low-level depressive or Low. Both of these clusters were quite stable after an initial higher center (Cluster #3) or after a center measured second in time that was slightly higher than the others (Cluster #2). The timing of the higher CES-D scores is what affected the slope making it negative. Thus, with the exception of the single high score, these three clusters taken together represented the majority of women in this sample (139 or 64%) who were generally stable, although at slightly different

levels. It could also be said that the majority of women either had CES-D scores in the non-depressive range or were improving (Cluster #7 and Cluster #3).

Cluster #1, the fourth most populous, was titled Moderate or Mod. Scores were generally stable. The positive slope could be attributed mostly to a low initial CES-D score. Four clusters taken together, #1, #2, #3, and #7, the large majority of the total sample (76%), had relatively stable scores over the 8 or 9 years, although at different levels, Non, Low, and Mod.

Cluster #6, nearly as populous as #1, was labeled Rise and Fall. While SLOPE99 was even slightly higher than in Cluster #1, the level of CES-D scores was lower, and the overall pattern was more quadratic in nature. Cluster #8 was described as Rising or Emerging depressive symptoms. This cluster showed fairly stable scores until the two measured last in time. As can be seen graphically from the grouped bar graphs of Figure 1, Cluster #8 was almost a mirror image of Cluster #3. Cluster #9, the least populous general cluster, was labeled Fall and Rise High, clustering those women with the highest scores. The quadratic slope of Cluster #9 was due to the higher variability of CES-D scores at high values, as seen by the general fan shape of data points in the scatterplot shown in Figure 2 of the variables MEAN99 by RANGE99. Cluster #9 was also noted to be the most heterogeneous within the cluster. This heterogeneity can be seen in a box plot of SLOPE99 (Figure 3), which indicated slopes were the most variable in Cluster #9 compared to other clusters. However, within Cluster #9, the scatterplot of MEAN99 by SLOPE99 (Figure 4) demonstrated that at least half of the women had slopes between  $-0.1$  to  $+0.1$ , again an indication of overall stability at a particular level, even though individual

women demonstrated somewhat high range and high standard at a stable level overall. The latter are shown in the boxplots of RANGE99 (Figure 5) and SDCESD99 (Figure 6).

A final review of all graphs and charts revealed that clusters #2, #3, #6 and #8 differed mostly around the trends over time, as seen in SLOPE99 (Figure 3). As mentioned previously, Clusters #2 and #3 had slightly higher scores earlier for a decreasing or negative slope, and Clusters #6 and #8 had slightly higher scores later for an increasing or positive slope. These four clusters otherwise tended to have centers near a CES-D score of 10, a level labeled Low.

Line graphs of each woman's longitudinal pattern of CES-D total scores were also created. Figures 7 through 15 have provided examples of the individual woman from each cluster with the smallest distance from the centers of the cluster. Reference lines at 16 and 24 have been added to represent the standard and higher cut-off scores for major depression. Indications of MTS have been added for visual association (or its lack). At the time the graphs were created an additional CES-D score, CESD00, had been added; however, the final score was not part of the data set submitted for k-cluster analysis. Because these are actual data and not graphs of the statistically constructed centers of clusters, the single phrases naming the cluster patterns did not necessarily match the visual shape of the line graph perfectly, especially for clusters #2, #3, and #8 (see Chapter 5).

The example of Cluster #2, labeled Low, has a slope that appears more negative, with a pattern that seems to be falling more than the example for Cluster #3. In spite of the fact that the example for Cluster #3, labeled Falling, had the smallest distance from

the cluster centers, and had a SLOPE99 center of -0.96, this individual woman's pattern overall is quite stable and flat. It was also evident that the most representative example of Cluster #8 apparently had an overall slightly negative or falling slope, even though Cluster #8 was labeled Rising Low or Emerging and had the largest absolute value (of all clusters) of SLOPE99 (see Table 6).

Inspection of line graphs from all the women in a cluster highlighted the fact that there were individual women whose pattern and/or calculated distance from the cluster centers was large. That is, these women seemed less typical of the cluster or of the values of the cluster centers. In addition, for some clusters, especially #9, there seemed to be large heterogeneity within the cluster. Visible differences in slopes from what the SLOPE99 centers would imply were a particularly surprising result; some women in cluster #9 had falling slopes and some had rising slopes.

A table of summary measures for each k-cluster has been provided in Table 6. Comparable information of summary measures from hierarchical agglomerative cluster results has been provided in Table 7 in the next section. In Table 6 it may be seen again that Clusters #2 and #3 have only little difference between them, and then it is mostly in RANGE99 and SLOPE99. Clusters #6 and #8 have even less difference between them in these summary measures, and again the main difference between Clusters #2 and #3 and Clusters #6 and #8 is in direction of SLOPE99. The clusters with the highest MEAN99 have the highest RANGE99, that is, Clusters #1 and #9.

### **K-Cluster Summary**

Iterative cluster analyses of longitudinal patterns of CES-D scores from 216 women providing at least six annual scores was conducted and a solution of 9 clusters was deemed best. These clusters were descriptively labeled Stable Moderate (Cluster #1), Stable Low (Cluster #2), Falling Low (Cluster #3), Rare-1 (Cluster #4), Rare-2 (Cluster #5), Rise & Fall (Cluster #6), Stable Non-depressed (Cluster #7), Rising Low (Cluster #8), and Fall & Rise High (Cluster #9). Four levels were thus identified, non-depressed, low, moderate, and high. Five longitudinal patterns were also identified, stable, falling, rising, rise & fall and its mirror pattern fall & rise.

The five patterns could be abstracted into two general patterns: stable and changing. Stable non-depressed-level scores was the single most common pattern (Cluster #7); there were also scores relatively stable at low (Cluster #2) and at moderate levels (Cluster #1). Patterns of change included scores falling (Cluster #3), rising (Cluster #8), and those having more than one direction of change within the years studied (Clusters #6 and #9). The clusters with very few women were dramatically changing, including extremely high scores at some time points (Clusters #4 and #5).

### **K-Cluster Discussion**

What was most valuable from all the k-cluster analyses that were attempted was the consistency in general cluster results. The largest cluster of non-depressed stable women was always apparent. A cluster with high centers either throughout all years or most years was another persistent result. Most k-cluster results also had at least two clusters that were changing over time, although the direction and consistency of change

were not reproduced exactly in different analyses. Because of the slight differences in results between different k-cluster analyses, population sizes and summary measures such as mean, standard deviation, etc. displayed in Table 6 cannot be taken as final or exact.

It would be difficult to attribute any clinical significance to the differences between clusters #2 and #3 and to the differences between #6 and #8. Again, the main differences between all four of these clusters was the timing of one or two higher yearly CES-D scores which in turn affected the direction and steepness of slope, the last being the main difference between all four clusters. Nevertheless, the importance of slope as one of the variables submitted for clustering was vital; it was one way to make cluster analysis applicable for not just a randomly ordered set of variables but for a longitudinal pattern of repeated measures.

The potential problem with iterative or k-cluster techniques centered around the inability to determine whether or not an optimum clustering solution was actually obtained (Aldenderfer & Blashfield, p. 48). Trying to address this problem led to the next set of cluster analyses.

## **HIERARCHICAL AGGLOMERATIVE CLUSTER ANALYSIS**

### **of SUMMARY MEASURES**

#### **Analytic Procedures**

#### **Type Of Cluster Analysis**

Hierarchical agglomerative clustering was the second approach used. The purpose of using this type of clustering was to be able to avoid the problem with iterative or k-cluster analysis of not knowing if the result produced was the optimal one.

### Number Of Cluster Solutions

Number of clusters specified was 3, 4, and 5, keeping all other parameters the same. The decision about number of clusters was based on the consideration that the number of solutions specified should be determined by theory or expectations underpinning the topic of the research, and by the number of variables submitted for clustering (Aldenderfer & Blashfield, 1984 p. 20). Because of the results from the k-cluster analyses five solutions might have been ideal (stable non, stable low/moderate, high, rising low/moderate and falling low/moderate). However, because only three variables were submitted for clustering the number of solutions would best be limited to no more than three, according to a statistical consultant (Robert Burr, personal communication).

### Selection Of Variables

Unfortunately, in hierarchical cluster analysis a case with a single missing data point is deleted entirely. Requiring six or more annually repeated CES-D scores would have reduced the sample size to 71 women, thereby losing a great deal of information that was available. Therefore, in order to capitalize on the available data set and maximize the sample size, summary measures only were used. Summary measures included a measure of level (MEAN99), overall change over time (SLOPE99), and some measure of variability such as RANGE99, SDCESD99, or CVCESD99 (see k-cluster section before). Use of only summary measures allowed inclusion of even those women with few annual scores, since they could all have summary measures calculated. Summary measures also

were expected to capture a better categorization of longitudinal patterns than the k-cluster analyses, particularly with regards to overall variability and non-stable patterns.

The distribution of the summary variables was inspected. Neither CES-D scores themselves, nor the summary measures were normally distributed, except for SLOPE99. Transformations were computed of the variables which were not normally distributed in order to follow standard statistical procedures, (Aldenderfer & Blashfield, 1984, p. 20). First, an adjusted or unstandardized residual of RANGE99, labeled RANGEADJ, was computed. All summary variables, MEAN99, SDCESD99, CVCESD99, SLOPE99, RANGE99 and RANGEADJ, were further transformed to a percent of their range, labeled PMEANCES, PSDCESD99, PCVCESD99, PSLOPE99, PRANGE99, AND PRANGEAD. This was, therefore, a second transformation of RANGE99. A Z-transformation, or standardization, of all variables submitted for clustering was a second transformation of the summary variables, and a third transformation of the RANGE99 variable. All sets of summary measures have been listed in Figure 16 of options for hierarchical agglomerative clustering.

#### Method Of Hierarchical Clustering

Ward's method was utilized first for hierarchical agglomerative cluster analyses. Analyses were also conducted using Complete (Furthest Neighbor) and Between-Groups Average (B-average) methods.

#### Measure Of Hierarchical Clustering

Squared Euclidean measure is obligated for Ward's method. Additional hierarchical analyses using both Complete and Between-Groups Average methods but

varying the measures were conducted. Other linkage measures or measures of similarity that were considered for use were City Block or Manhattan, and Chebychev.

Figure 16 has displayed the algorithm that was used for each option of choosing parameters for the hierarchical cluster analyses presented here. It can be seen that 4 sets of variables, standardizing Z transformation, 3 numbers of solutions, 3 methods and 3 measures were potentially specified (although Ward's method is limited to the one measure, Squared Euclidean). A combination of all of these possibilities would have required 168 different analyses. All 168 were not conducted, however. As improvements in the results began to be appreciated, combinations of choices for the specifics of subsequent cluster analyses were limited to those that demonstrated the greatest distinction between clusters without overlap.

For ease of reporting, a standard format was adopted. The Method was listed first. The measure was listed next unless Ward's method was used, in which case the measure was always Squared Euclidean and therefore occasionally omitted. The number of cluster solutions specified was third. Fourth was the variable used as a measure of variability. Last was whether a standardized Z-transformation was applied to all variables before submitting them for clustering. An example would be: Complete City Block 4-Cl Prangead Z.

## Results

### Sample

Standard deviations of summary measure variables were large and pattern determination within clusters was deemed unreliable when the sample included women

with only three or four annual CES-D scores. Requiring five or more annual CES-D scores resulted in a reasonable improvement while maintaining the goal of maximizing sample size. There were 249 women from the SMWHS who had at least 5 annual CES-D scores.

#### Number Of Cluster Solutions

Four solutions were specified in the analysis judged best. Four was a reasonable compromise between three and five, and there was good definition of and discrimination between clusters compared to three or five cluster solutions. Increasing the number of cluster solutions from three to four maintained two of the clusters exactly and split the third cluster into two, primarily splitting out nine women who could have been labeled as outliers within the third cluster. This “new” fourth cluster of nine women was distinguished by much higher values of MEAN99, RANGE99, and (positive) SLOPE99. By increasing the specified number of solutions to five, again the only change was that one of the four clusters was split into two, by splitting out four women who could have been labeled as outliers within the fourth cluster. This “new” fifth cluster of four women was distinguished mostly by SLOPE99, including the woman with the greatest negative SLOPE99 of  $-4.97$ . Since increasing numbers of solutions specified tended to separate outliers of smaller and smaller numbers of women into their own cluster, further increases were not thought to be helpful in understanding general longitudinal patterns.

#### Selection Of Variables

A standardized Z-transformation of the summary variables MEAN99, RANGE99, and SLOPE99 provided the hierarchical cluster result judged best in distinguishing

between clusters with least overlap. Use of the P-transformed variables produced clustering with greater overlap than the Z-transformation, particularly visible in scatterplots and boxplots of the same sets of variables. Also, using SDCESD99 or CVCESD99 produced greater overlap between clusters than using RANGE99.

#### Method And Measure

Ward's method and obligatorily Squared Euclidean measure were used in the 4-cluster hierarchical analysis judged as best. This method and measure again discriminated the best between clusters, with the least overlap between them. City Block or Manhattan and Chebychev measures did not improve the discriminatory power of clusters produced by Complete or Between-Groups Average methods. The latter two methods were significantly inferior to Wards method in discriminating between distinct clusters even when keeping constant the use of Squared Euclidean measures.

#### Clusters

The satisfactory discrimination between clusters can be seen visually in the bar graph of the yearly means of CES-D scores in Figure 17. Cluster #1 was a stable low cluster. Cluster #2 was a moderate level cluster with a falling slope. Cluster #3 was a low-to-moderate level cluster with a rising slope. Cluster #4 was a high cluster with a rising slope. RANGE99 was also distinguished between clusters, as seen in the box plot of Figure 18. While Clusters #2 and #3 might look similar in RANGE99, the difference between the two was actually demonstrated in SLOPE99 (Figure 19). It can be readily seen that the main difference between Clusters #2 (Falling Moderate) and #3 (Rising

Low) was the direction of SLOPE99, a conclusion that could also be derived from the grouped bar graph of the yearly means shown in Figure 17.

The main benefit of this particular analysis, however, was demonstrated in scatterplots, where the overlap between clusters was minimal. Cluster #1 (Stable Low) was clearly defined as lowest MEAN99 and lowest RANGE99 (Figure 20). Cluster #4 (High) was more widely scattered in the plot but had some of the highest RANGE99 in combination with the highest MEAN99. The rather narrow diagonal pattern of the data points for Cluster #4 indicated that increasing MEAN99 was associated with increasing RANGE99. Clusters #2 (Falling Moderate) and #3 (Rising Low) once again showed the greatest overlap, with #2 being the most widely spread over both MEAN99 and RANGE99.

Figure 21 of a scatterplot of MEAN99 by SLOPE99 again showed a fairly clear discrimination between clusters. Cluster #1 (Stable Low) was not only clustered most tightly and had the lowest MEAN99 (and lowest RANGE99 from the previous plot) but also had the least trend in SLOPE99, bunched narrowly around the value of zero. Slopes tightly clustered around zero would be consistent statistically with a low range combined with a low mean. In addition, Cluster #4 (High) was clearly distinguished by some of the highest values of positive SLOPE99 and the highest half of the graph for MEAN99. Clusters #2 (Falling Moderate) and #3 (Rising Low) here were better distinguished, #2 clearly having negative and #3 positive slopes.

The scatterplot of SLOPE99 by RANGE99 (Figure 22) once again demonstrated a clear discrimination between clusters. Here it may be seen that Cluster #3 (Rising Low)

had a somewhat lower range than #2 (Falling Moderate) and definitely lower range than #4 (High), while almost sharing the same positive slopes of Cluster #4. Summary measures and population sizes of each cluster have been provided in Table 7.

### **Hierarchical Cluster Summary**

Hierarchical cluster analyses generally replicated the findings of k-cluster analyses. Once again a large cluster with non-depressed and low MEAN99, low RANGE99, and stable (zero) SLOPE99 was consistently the single largest hierarchical cluster. Other clusters had a moderate level MEAN99 and a moderate RANGE99; they were distinguished, however, by positive and negative SLOPE99. There was generally a smallest cluster, with the highest MEAN99, RANGE99 and SLOPE99.

### **Hierarchical Cluster Discussion**

Hierarchical clustering provided a condensed summary of general patterns, easy to interpret and remember. In particular, hierarchical cluster findings included recognizing the importance of slope or general trend over time in the yearly CES-D scores, as well as level. Three of the four hierarchical clusters demonstrated a trend over time in longitudinal pattern, and SLOPE99 was the primary distinguishing factor between Clusters #2 and #3.

Limitations of hierarchical clustering included the inability, even with increasing numbers of clusters specified, to distinguish between a non-depressed stable and a low stable longitudinal pattern. While the 9-cluster solution of k-clustering produced some possibly redundant and/or outlier longitudinal patterns, the 4-cluster solution of hierarchical clustering on the other hand might not have distinguished between some

meaningfully different longitudinal patterns, especially non-depressed and low level of symptoms.

There were also general limitations in the cluster analyses related to variables selected for clustering. Although the slope of a simple regressed line was used (SLOPE99), there were generally extremely low values of  $R^2$ , a measure of the fit of the line to the data. The mean average  $R^2$  was .21 with a standard deviation of the  $R^2$  statistic of .20. The  $R^2$  of a polynomial curve fitted to the data points was much higher than that of a simple line. Unfortunately, for a maximum of eight data points through CESD98, the majority (65%) of women's CES-D scores required a 5<sup>th</sup> or 6<sup>th</sup> order polynomial to achieve an  $R^2$  of at least 0.80. Additionally, the fitted curve usually included impossible values below zero. Because of the difficulty interpreting a 6<sup>th</sup> order polynomial for eight data points, a simple line was judged to be more useful.

In another limitation, SLOPE99 was markedly affected by a single high (or low) CES-D score depending on where it fell in the time sequence of CES-D scores. If a high CES-D score occurred early in the time sequence the slope was negative and conversely if a high score occurred late the slope was positive. A single high score in the middle of the time sequence tended to flatten or level the slope towards zero. Yet since the timing of the six or more years of CES-D scores was arbitrary within midlife, little can be concluded regarding the timing of a single high score.

Another limitation lay in the capabilities of cluster analysis in general. By definition and design, cluster analysis does not provide a complete review of all the patterns that might be identified in a data set. In fact, cluster analysis is not specifically

designed for identification of longitudinal patterns, although it was serendipitously quite informative. Nevertheless, inspection of examples of actual longitudinal patterns, such as was used for establishing reliability of cluster results, revealed that there was further understanding to be gained from a more thorough and systematic approach to inspection. Phenomenological analysis would be able to provide such an approach (see Chapter 5).

### Validity And Reliability

While Aldenderfer and Blashfield recommended that some validation procedure always be performed and reported for cluster analysis, they specifically reviewed only Monte Carlo techniques (Aldenderfer & Blashfield, 1984, p. 74, 81). Monte Carlo techniques use

random number generators, to create a data set with general characteristics matching the overall characteristics of the original data but containing no clusters. The same clustering methods are used on both the real data and the artificial data, and the resulting solutions are compared (p. 67).

It was considered for this set of analyses, however, that repeating many cluster analyses, as performed here, served the same purpose. The many analyses used real data as well. The general consistency of results, regardless of type of clustering and specific variables submitted and regardless of the variations in number of cluster solutions and methods and measures used for hierarchical analyses, was probably as useful in establishing validity and reliability as Monte Carlo techniques. In addition, in order to evaluate the results of early k-cluster analyses a non-mathematical approach was also used. Visual inspection of longitudinal graphs from individual women in a cluster sometimes revealed aspects of the longitudinal pattern within a particular cluster that were not very consistent. As

mentioned earlier, the k-cluster analyses were evaluated and refined to include SLOPE99 by visual review of line graphs.

In summary, both types of cluster analysis, iterative and hierarchical, had similarities, and each provided a special contribution to understanding longitudinal patterns of depressive symptoms in midlife women. Iterative clustering allowed a refined discrimination between patterns among groups of women, particularly distinguishing non-depressed, low, moderate, and high levels, and distinguishing changes in slope within an individual woman's pattern (rise and fall, fall and rise). Hierarchical clustering dramatized the value of including slope, that is, the general trend over time, in analyzing longitudinal patterns.

The set of both types of cluster analyses elucidated the parameters of longitudinal patterns of CES-D scores that varied across midlife women. These parameters were a woman's mean score over the years, her range of scores over the years, the slope of a fitted line to her scores graphed across time, and changes in these three parameters over time. These parameters may be interpreted clinically as the average level of depressive symptoms, lability of depressive symptoms over time, the general trend over time of changes or stability in depressive symptoms, and frequency of changes in level, lability, and trend, respectively.

### **PREDICTORS AND CONFOUNDERS**

In order to address the second aim of the study, the relationship of depressive symptoms to selected predictors was examined. Predictors investigated were MTS and history of harassment, abuse and/or assault. Antidepressant drug use was examined as a

confounder. MTS (see Chapters 2 and 3) was coded as eight categorical (dummy) variables: pretransition, early transition, middle transition, late transition, postmenopause, HRT use, OC use, and other. One example, in 1999, of a distribution of women in the MTS categories, with the mean and standard deviation of total CES-D scores in each MTS, has been provided in Table 8.

#### Menopausal Transition Stage (MTS)

Nine two-tailed analyses of variance (ANOVA) were conducted, one for each year, where CES-D scores were compared by MTS. Results for each year have been displayed in Table 9. Significance was not adjusted for multiple comparisons; however, since multiple comparisons bias towards finding at least one spuriously significant statistic, non-significant findings are considered reliable (Kleinbaum et al., 1988, p. 362). The same sample was used as for the k-cluster analysis of 216 women with at least 6 CES-D scores longitudinally through 1999.

The exact number of women who had concomitant MTS data each year varied slightly, from a low of 109 for CESD95 to a high of 205 for CESDt0 at entry to the study. No significant differences in CES-D scores were found by MTS in any year. F-statistics ranged from .3770 in 1996 to 1.611 in 1998. Significances of the F statistics ranged from a low of .081 in 1998 to a high of .977 in 1996, with slightly varying degrees of freedom depending on the year's sample size and number of MTS categories actually having women assigned to them.

A second set of nine two-tailed t-tests, again one for each year from the same sample of 216 women used for the k-cluster analysis, compared CES-D scores between

two groups, late transition and all other stages combined. Late transition was chosen because it has been implicated in the literature as the most likely time for an increase in depression (Avis et al., 1994; Dennerstein et al., 1997; Kaufert et al., 1992). Results have been presented in Table 10. Again without any adjustments for multiple comparisons needed, there were no significant differences between total CES-D scores of women in late transition compared to all other women as a whole. Analyses were repeated every year using middle transition with the same results; there were no significant differences any year between women in middle transition compared to all other women. The same comments apply about multiple comparisons.

#### History of Assault/Abuse

History of childhood maltreatment, abuse and/or neglect, and history of adult harassment, assault and/or rape have also been proposed as predictors of adult depression (see Chapter 2). In the 1995 annual questionnaire one item asked about sexual harassment during the previous year, two items asked if there were sexual abuse and physical abuse at any time in the past, and three yes/no items asked whether the sexual and/or physical abuse occurred as a child, teenager, and/or adult (see Appendix A). However, numbers in each category were small because the 1995 questionnaire was not administered to almost half of the women (see Chapter 3). In order to obtain information from a more complete sample the 1999 annual questionnaire contained a single yes/no item that asked if the respondent ever had been sexually assaulted, abused, or molested. Additionally the 1997 and 1999 annual questionnaires each asked three yes/no items, one whether the respondent has been sexual harassed, by her own definition, during the

previous year, one whether the respondent had been physically abused in the past year, and one whether she had been sexually abused in the past year. See Appendix A for wording of each item and descriptive statistics from the total SMWHS sample. A cumulative variable was also created for the k-cluster sample of 216 women to represent any history of any harassment, abuse or assault on any item on any questionnaire, coded as yes=1, otherwise=0.

Three  $\chi^2$  (Chi-square) analyses of each of the three individual abuse items from 1995 by k-cluster were not significant, using a two-tailed significance level  $\alpha=.05$ . However, more than one-fifth of the cells contained expected values less than five, and at least one cell contained an expected value less than one, thereby not meeting the conditions required for  $\chi^2$  analysis (Rosner, 1986, p. 340). A  $\chi^2$  analysis of the cumulative variable was also not significant, and again more than one-fifth of the cells contained expected values less than five and one cell had an expected value less than one. When repeated without the rare Clusters #4 and #5, conditions for  $\chi^2$  analysis were met. Nevertheless,  $\chi^2$  analysis of the cumulative variable by k-cluster was not significant at a two-tailed level of  $\alpha=.05$  ( $n=212$ ; Pearson  $\chi^2=9.093$ ;  $df=6$ ;  $p=.168$ ).

Only 174 of the 216 women from the k-cluster 9-solution results provided information on the 1999 history of sexual assault/abuse item. Again eliminating the two rare clusters, a  $\chi^2$  analysis of the remaining seven clusters showed a significant difference between clusters in proportion of women answering yes to the history of sexual abuse/assault item (Pearson  $\chi^2 = 15.23$ ;  $df=6$ ;  $p=.019$ ). Percentages and adjusted residuals for all clusters have been provided in Table 11. Cluster #9, the cluster with the

highest chronic CES-D scores, had the highest proportion of women who reported yes to sexual assault/abuse history, 75%. Cluster #8, rising low, had the lowest percentage reporting any history of sexual assault/abuse, 0%. The large stable non-depressed Cluster #7 had the second lowest proportion of abuse history, 23%. Other clusters with low and moderate level mean total CES-D scores had intermediate proportions from 30-37% of respondents reporting a positive sexual assault/abuse history. Inspection of adjusted residuals revealed that the two clusters, #8 Rising Low and #9 Fall&Rise High, were the ones significantly different from expected values, reading the residuals “roughly as z scores ... [where] values well below -2 or above +2...identify cells that depart markedly from the model of independence” (“Using crosstabs,” 1997, p. 71).

#### Antidepressant Drug Use

Longitudinal patterns of depressive symptoms cannot be fully understood without considering confounders such as interventions to treat symptoms. Effective treatment would be expected to lower CES-D scores and change longitudinal patterns by shortening depressive episodes. Data regarding use of treatment such as cognitive behavioral therapy (CBT), other forms of psychotherapy, or use of herbals such as Hypericum (St. John’s Wort) were not available. However, data about prescriptive antidepressant drug use concomitant with the CES-D score were available every year except CESD36 (see Chapter 3). For each of the other eight years  $\chi^2$  statistics were calculated for differences in proportions of women using antidepressant drugs between k-clusters. However, even excluding the rare clusters every year there remained too many cells with expected rates less than one for the  $\chi^2$  statistics to be interpretable (Rosner, 1986, p. 340). Simple

inspection of percentages was conducted, however. Excluding the rare clusters, most years the two clusters of highest scores (#1 and #9, stable moderate and fall & rise high, respectively) had the highest percentages of antidepressant drug users. The cluster with the lowest scores (#7, stable non-depressed) usually had the lowest percentage. Thus in general, antidepressant drug use was not a confounder but was more likely in women with high CES-D scores.

Changes in the total sample percentage using antidepressant drugs was also of interest. Table 12 has shown the historic trend toward increasing use of prescriptive antidepressant drugs in general among the same women, a fixed sample of 201 women who provided CES-D data for the cluster analyses and also provided data on antidepressant drug use. The percentage of the total sample using antidepressant drugs increased nearly four fold over 7-8 years. This increase might be due to an increasing prevalence over time of depression in the sample. However, based on anecdotal information it seems more likely to have been due to an increase in prescribing styles by community providers.

#### Predictors And Confounders Summary

MTS was not significantly related to cluster. The lack of significance was consistent, using stage as a dichotomous variable comparing late transition to all other stages, comparing middle transition to all other stages, as well as using stage as an 8-part variable comparing between pre-, early, middle, late transition, postmenopause, HRT users, and others. A history of sexual assault/abuse was significantly related to cluster. As expected, this history was significantly more likely in the cluster with chronic high

CES-D scores. Clusters with moderate and some with chronic low level scores had intermediate percentages of women with sexual assault/abuse histories, and the stable non-depressed cluster had the second lowest percentage of women with such histories. Unexpectedly, the lowest proportion, 0%, was in Cluster #8, the rising low cluster. Antidepressant drug use was highest every year in the cluster with chronic high scores, Cluster #9.

### Predictors And Confounders Discussion

The lack of relationship between MTS and longitudinal pattern of depressive symptoms, or even between a single stage such as late, or middle, and all others combined, was consistent with the majority of longitudinal studies of depressive symptoms in midlife women (see Chapter 2). The significant differences between clusters in proportion of women revealing history of sexual assault/abuse was consistent with other studies and their reviews. It also was consistent with the concept that chronic depression in adult women is often a long-term sequela of assault/abuse (Hall, et al. 1993; Johnson et al., 1999; Walker et al., 1999; Walker et al., 1992; Widom, 1998). The lack of significance of the  $\chi^2$  analysis for the composite variable (the history of any kind of harassment, abuse, or assault) was probably due to the rather consistently modest percentages across all clusters in history of such harassment and violence.

History of abuse has been surprisingly missing from many causal models and certainly most research investigating longitudinal patterns of depression (see Chapter 2). The overall prevalence of sexual assault/abuse histories in midlife women in the SMWHS, 32.5% of the sample of 194 women in 1999, was not inconsistent with other

studies (Hall, et al. 1993; Walker, Torkelson, Katon, & Koss, 1993). Even the stable non-depressed cluster (#7) had nearly one quarter of the women reporting a history of sexual abuse/assault. The explanations for why a minority of women with histories of abuse demonstrated resilience without adult depression has not been well investigated (Liem et al., 1997).

An unexpected finding was the significant difference in predictors between different longitudinal patterns. This difference implied that there were different causal phenomena underlying different longitudinal patterns of depressive symptoms. Sexual abuse might be predictive of more chronic and higher levels of depressive symptoms while another factor(s), such as new difficulties with resources or relationships (Kaufert, et al., 1992), might be more predictive of other longitudinal patterns (Rising Low, Cluster #8). Hays, Blazer, and Gold (1993) found in a study of elders that a rise in CES-D scores was less predicted by usual risk factors for depression, and was instead predicted by consistent high scores across two time points. Cowdry, Gardner, O'Leary, Leibenluft, and Rubinow (1991) also found different patterns of mood ratings in women with MDD, borderline personality, PMS, and normal subjects. Sheehan et al. (1995) have suggested that using subscales from a four-factor structure of the CES-D might provide an avenue for investigating differences in patterns of change over time.

Interpretation of the observation that antidepressant drug use was higher in the cluster with chronic high CES-D scores was puzzling. Similar correlation between concurrent antidepressant therapy and high CES-D scores was also found by Hall et al. (1993). It is possible that the CES-D scores would have been even higher without the

drug therapy. Nevertheless, the expectation that antidepressant drug use would be a confounder of low CES-D scores was not realized. Because the relationship of the potential confounder of antidepressant drug use with CES-D scores was in the opposite direction of that expected, it was not considered necessary to make adjustments for antidepressant drug use in the ANOVAs, t-tests, or  $\chi^2$  analyses.

The analyses using MTS and antidepressant drug use could be considered preliminary because both the dependent variable, total CES-D score, and independent variables varied across time simultaneously, requiring repeated cross-sectional analyses for every measurement point given the statistical methods used. Statistical approaches that accommodate such models, such as hierarchical modeling (Bryk & Raudenbush, 1992), have been planned because they can appropriately include time-varying predictors and confounders such as MTS and antidepressant drug use in statistical modeling.

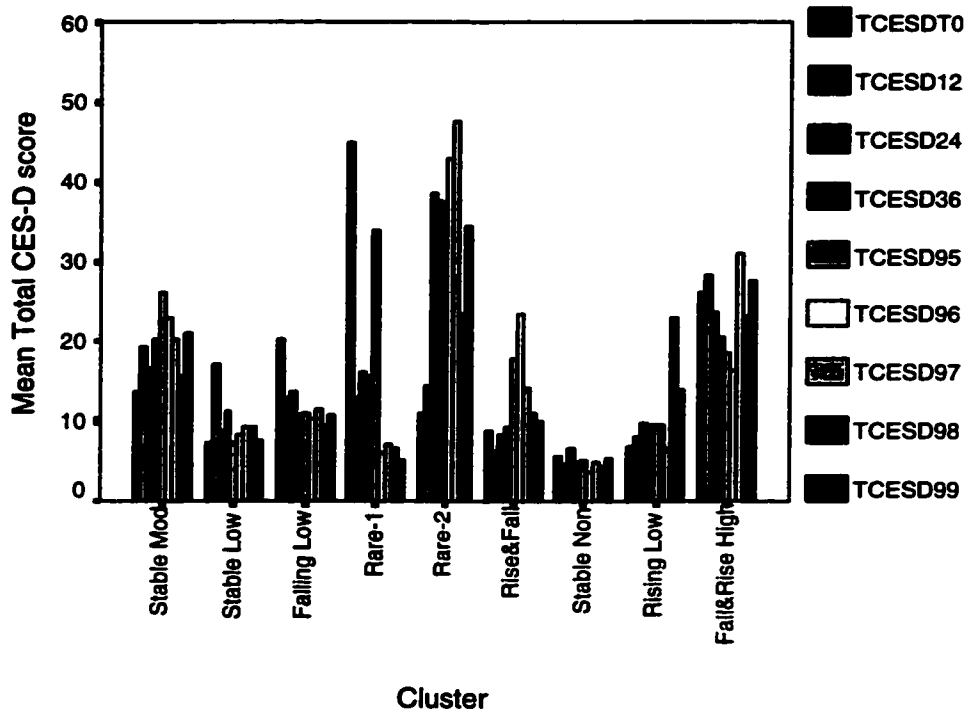


Figure 1  
Mean Yearly Total CES-D Score By K-Cluster

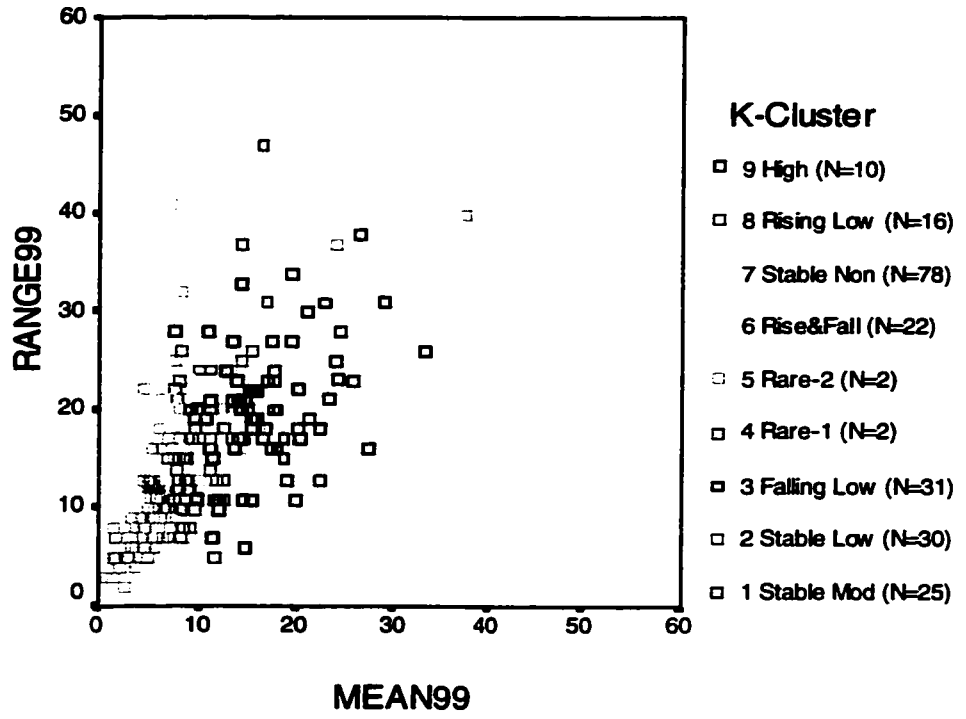


Figure 2  
Scatterplot of MEAN99 by RANGE99 by K-Cluster

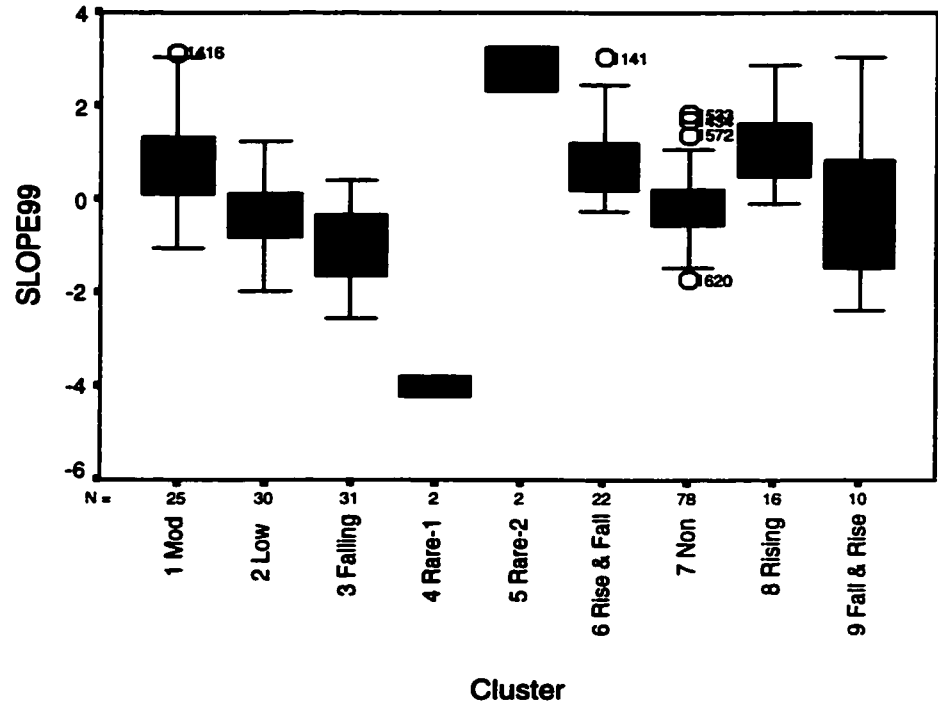


Figure 3  
Boxplot of SLOPE99 by K-Cluster

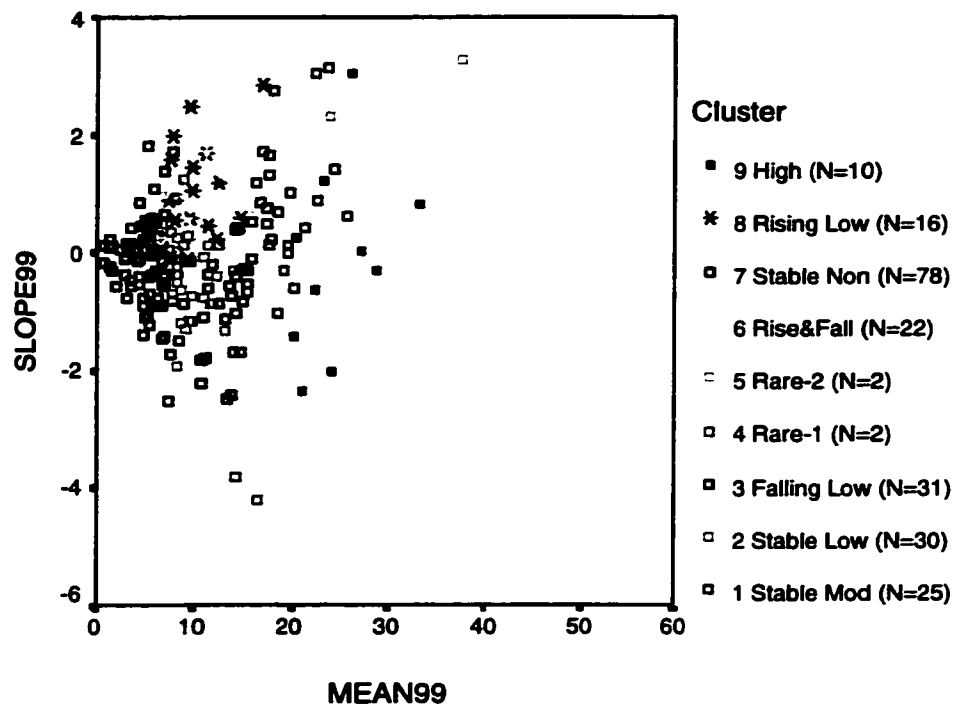


Figure 4  
Scatterplot of MEAN99 by SLOPE99 by K-Cluster

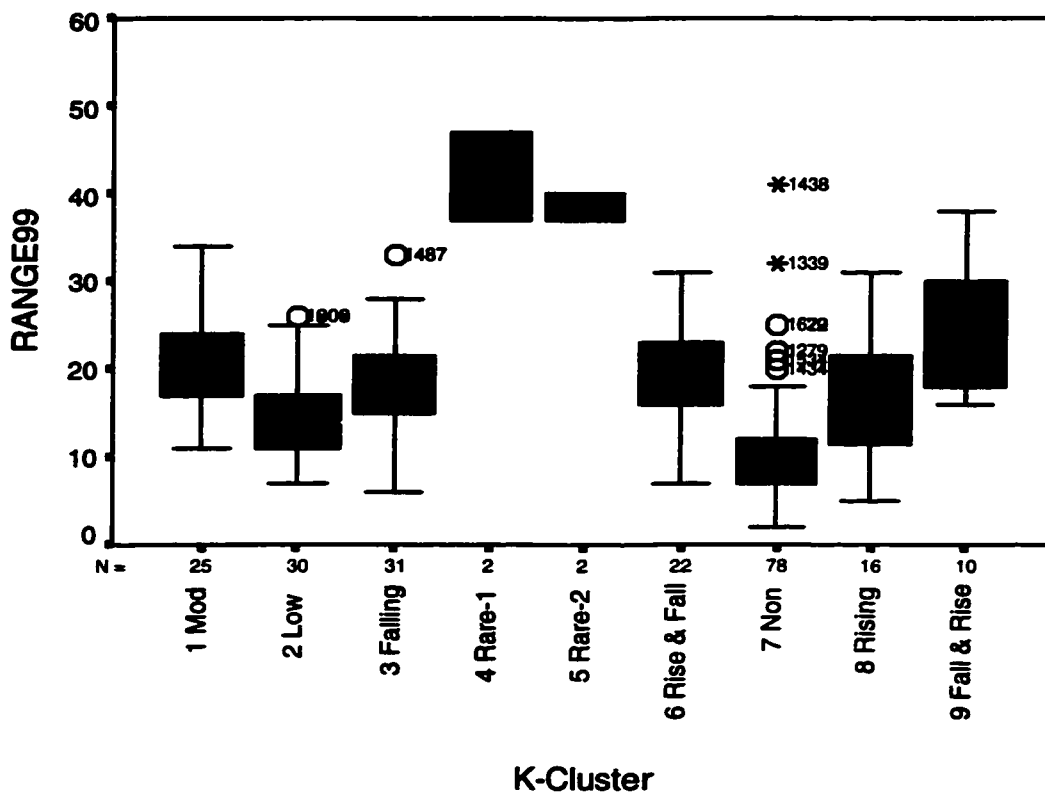


Figure 5  
 Boxplot of RANGE99 by K-Cluster

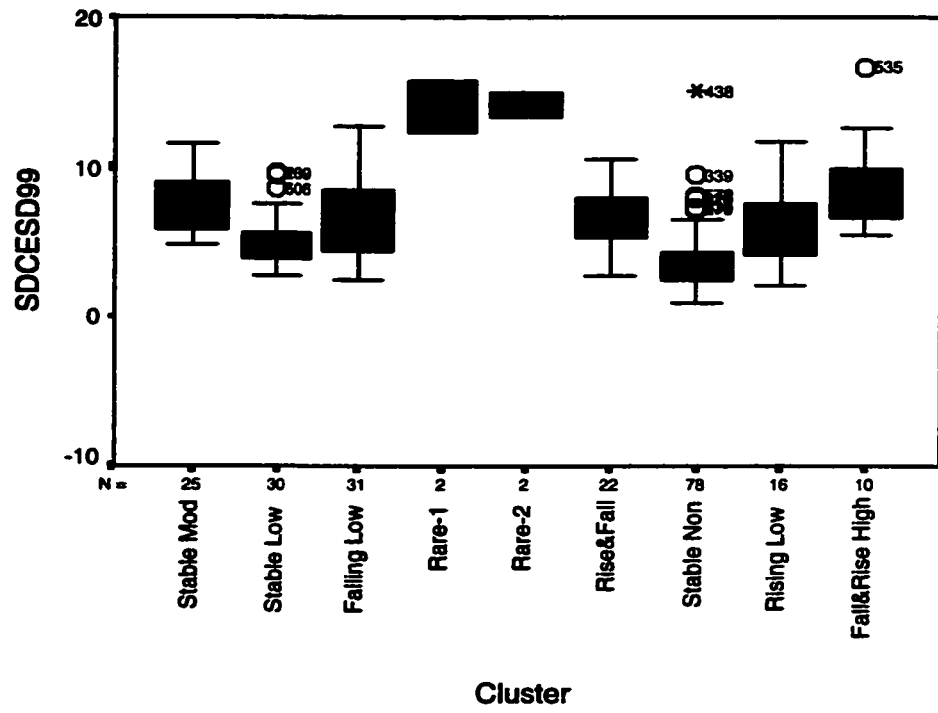
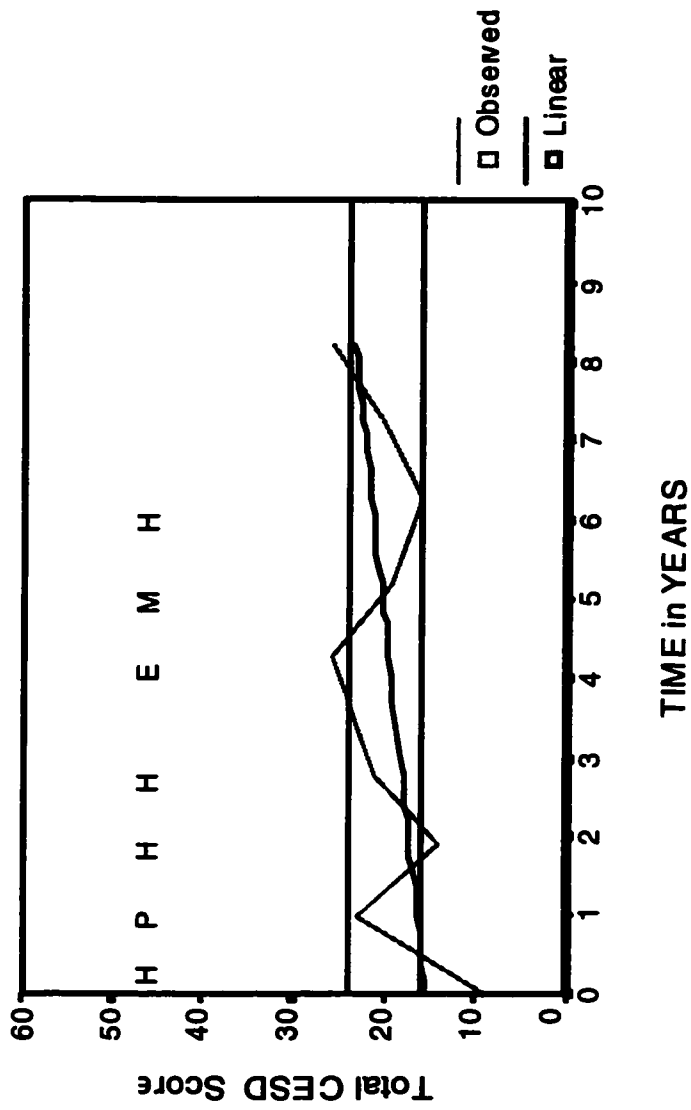
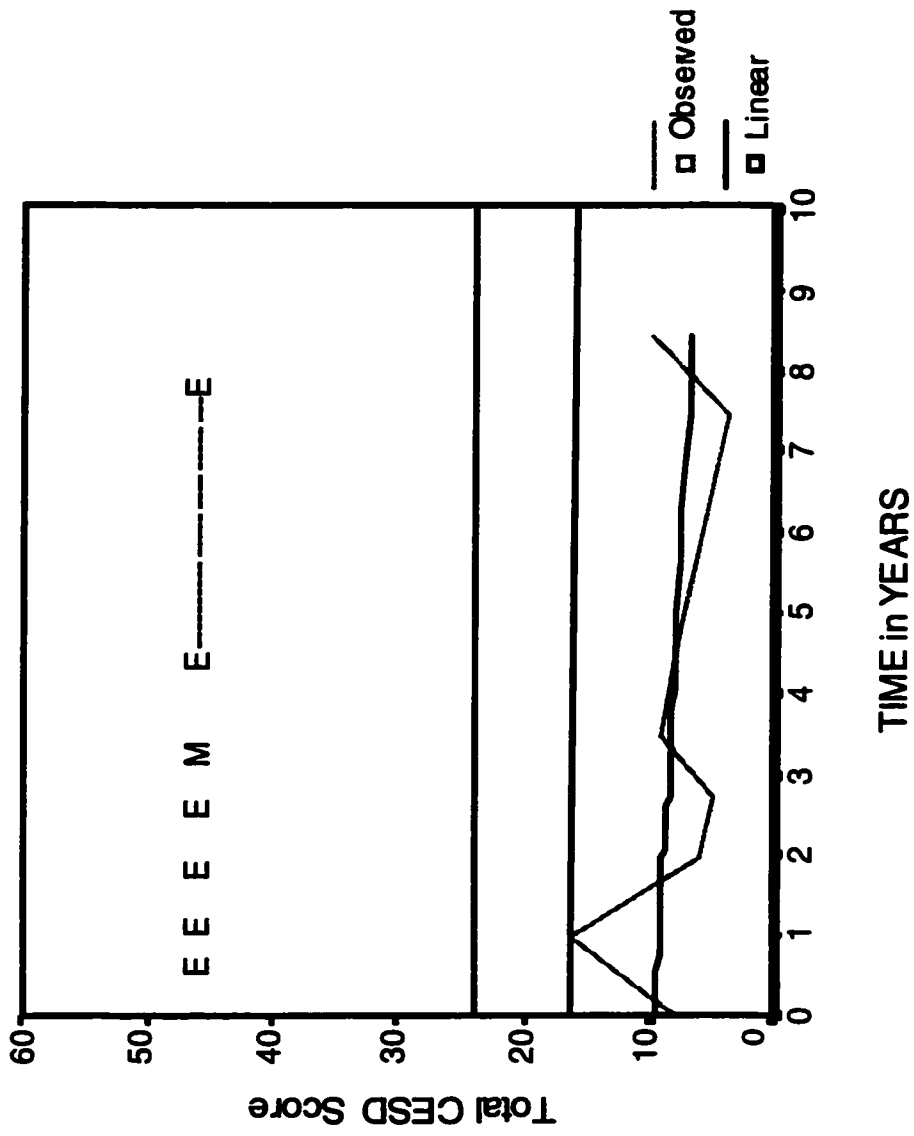


Figure 6  
Boxplot of SDCESD by K-Cluster



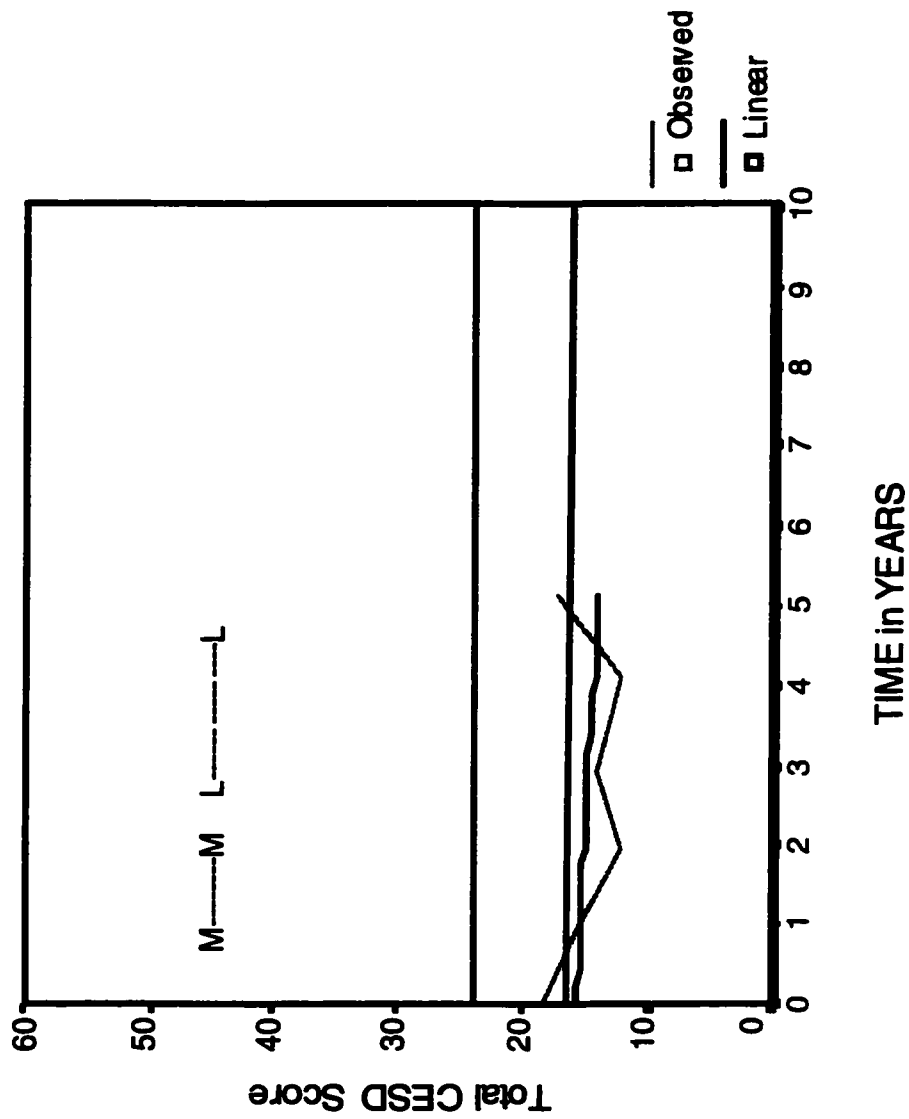
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT or OC use; O=Other; U=Unknown

Figure 7  
Example of K-Cluster 1 Stable Moderate



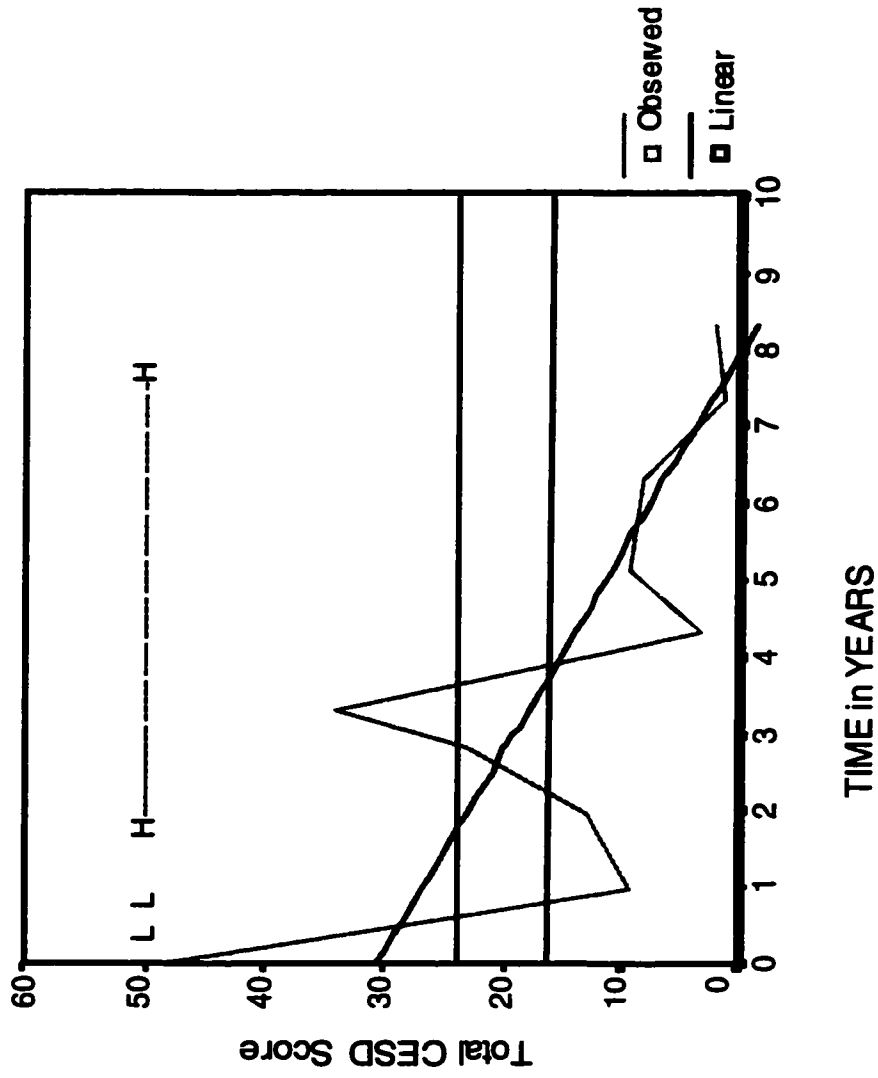
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT or OC use; O=Other; U=Unknown

Figure 8  
Example of K-Cluster 2 Stable Low



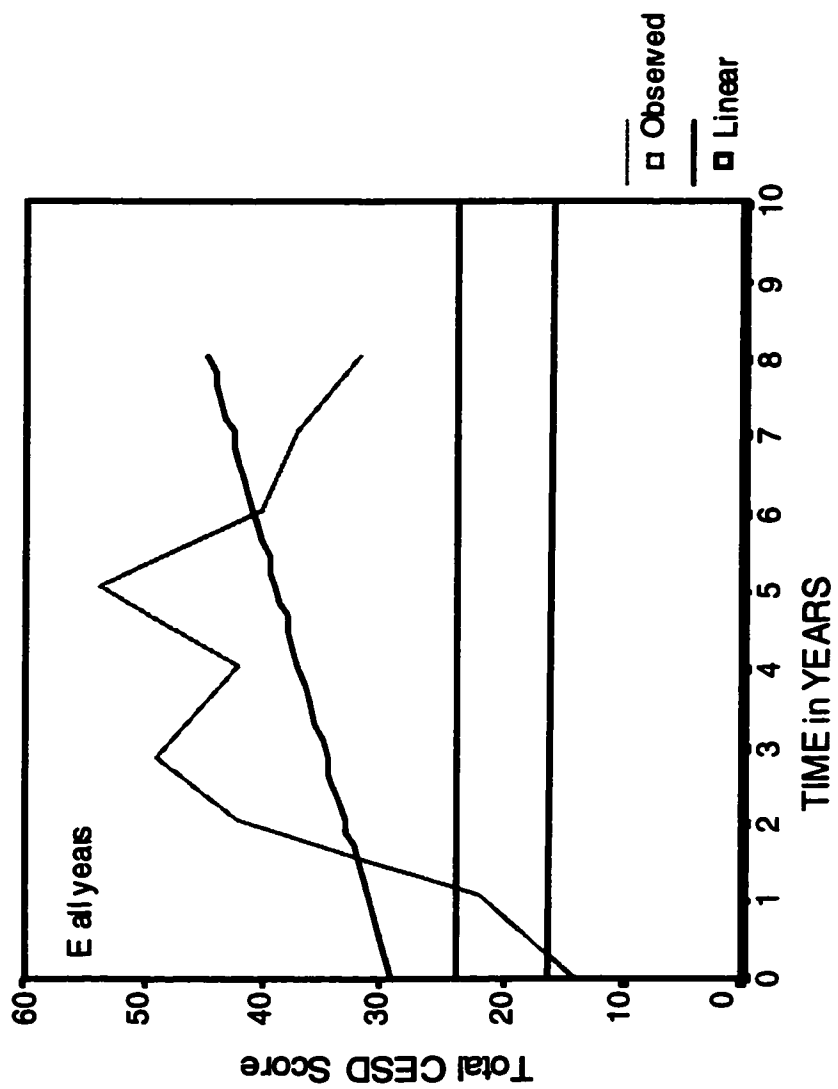
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT or OC use; O=Other; U=Unknown

Figure 9  
Example of K-Cluster 3 Falling Low



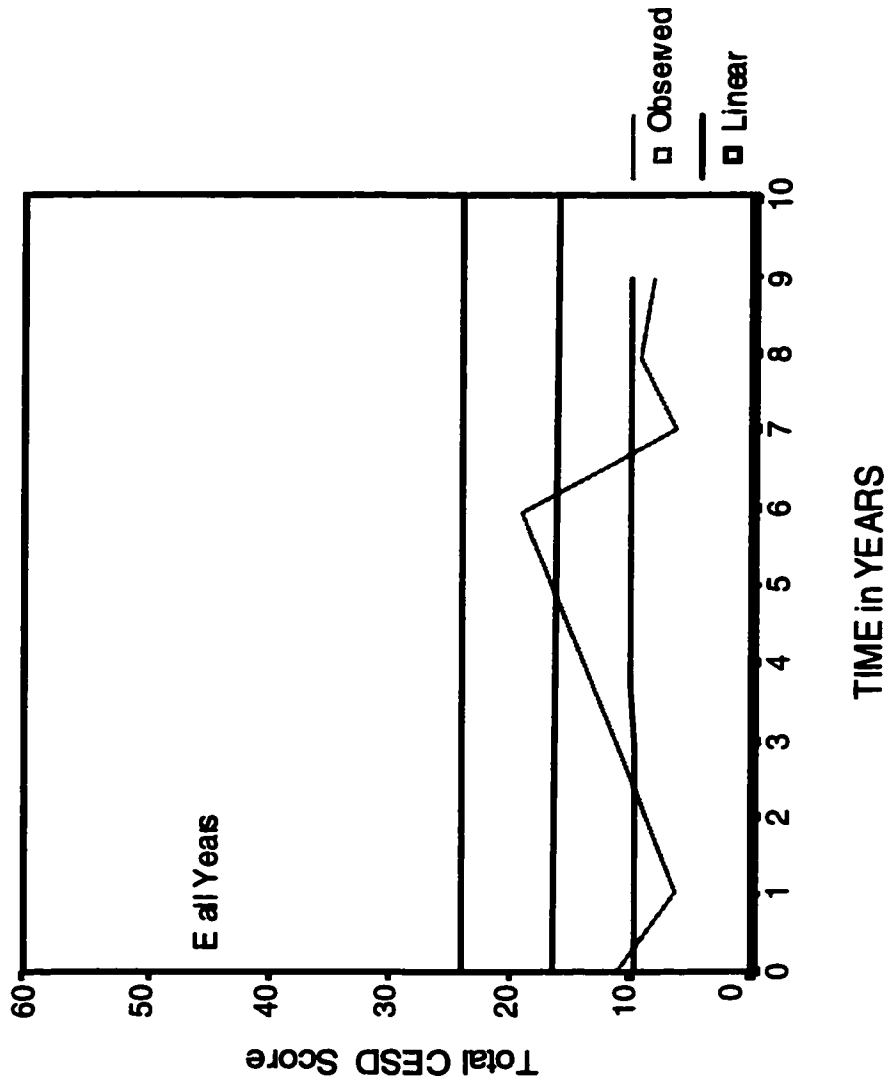
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT or OC use; O=Other; U=Unknown

Figure 10  
Example of K-Cluster 4 Rare-1



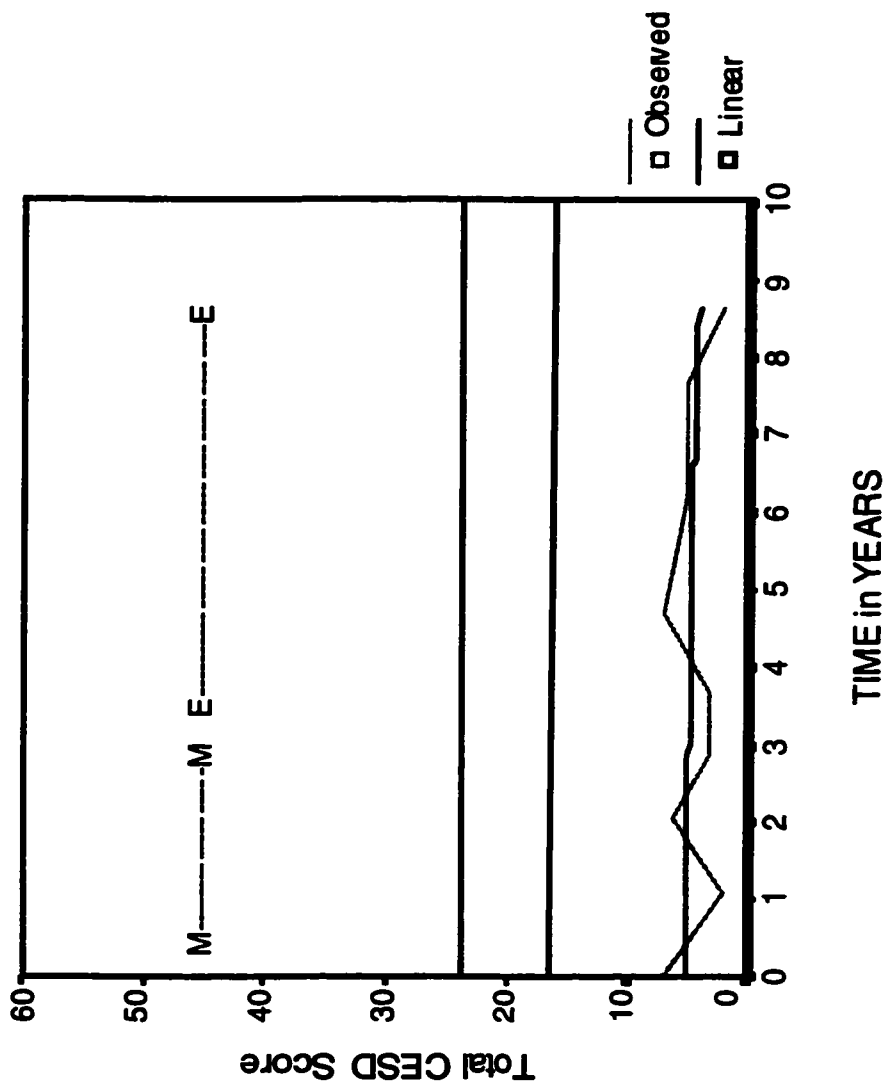
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT or OC use; O=Other; U=Unknown

Figure 11  
Example of K-Cluster 5 Rare-2



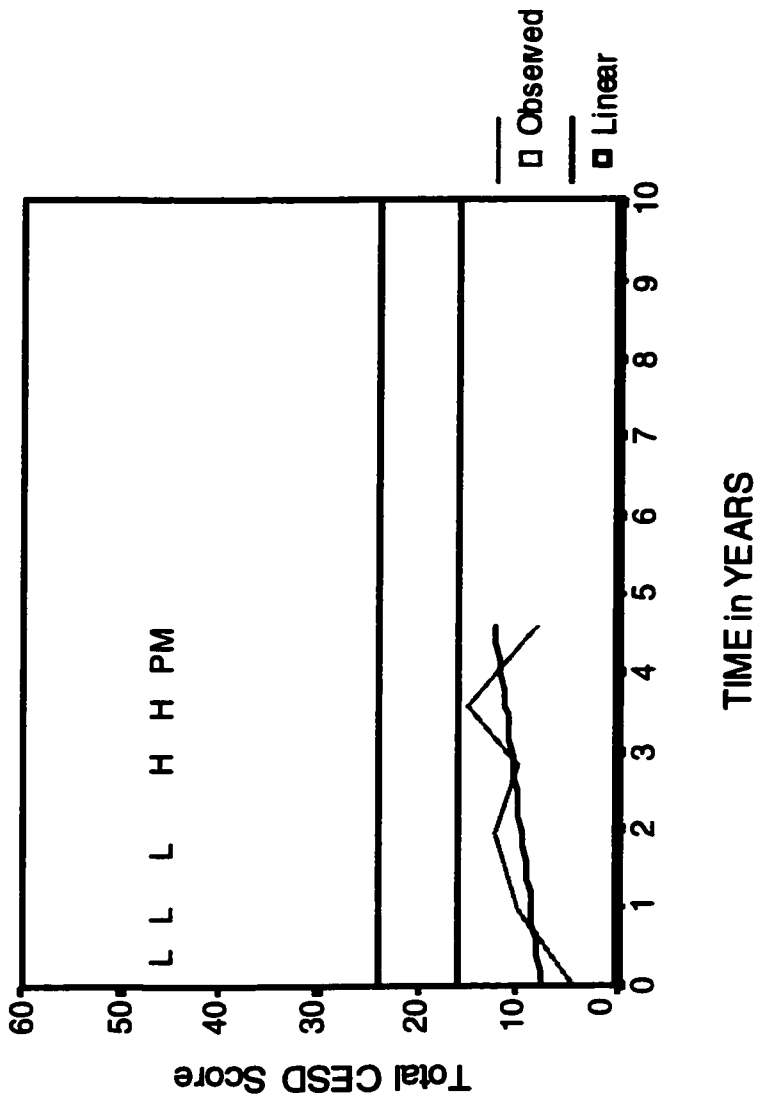
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT or OC use; O=Other; U=Unknown

Figure 12  
Example of K-Cluster 6 Rise&Fall



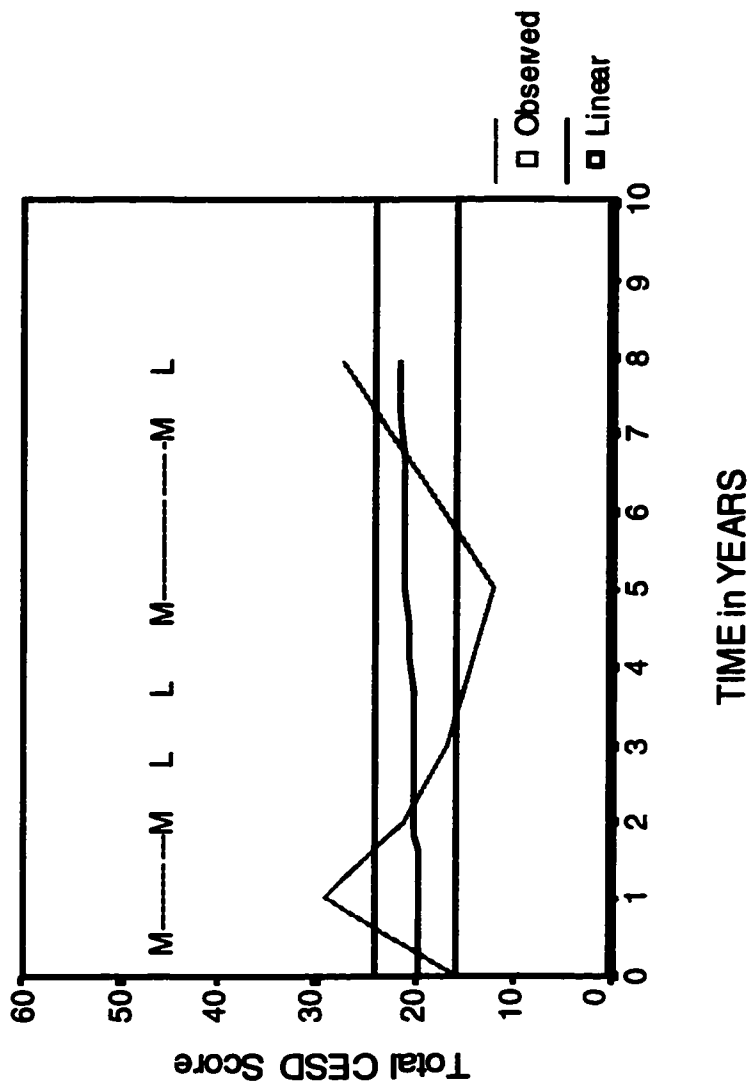
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT or OC use; O=Other; U=Unknown

Figure 13  
Example of K-Cluster 7 Stable Non-Depressed



P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT or OC use; O=Other; U=Unknown

Figure 14  
Example of K-Cluster 8 Rising Low



P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT or OC use; O=Other; U=Unknown

Figure 15  
Example of K-Cluster 9 Fall&Rise High

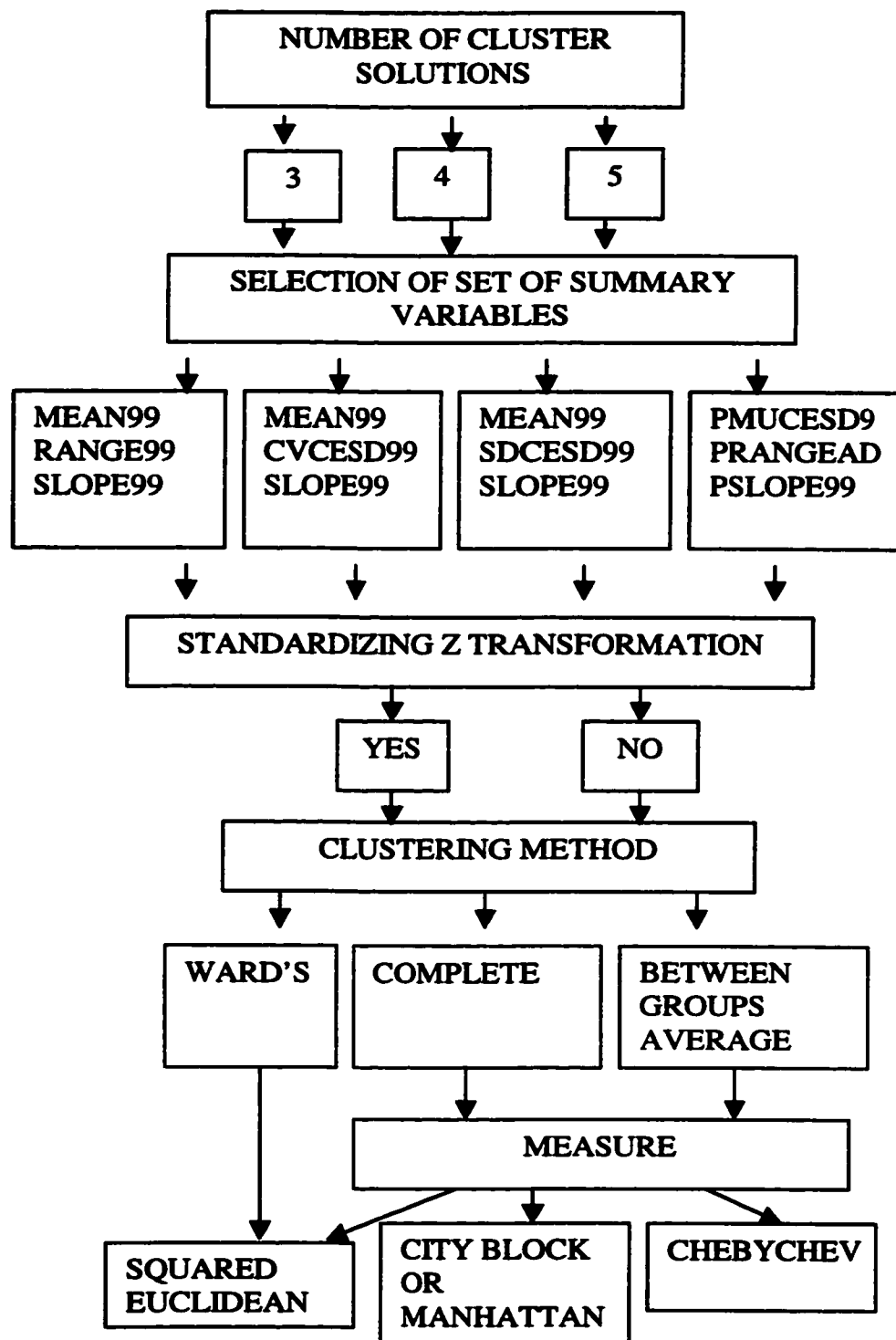
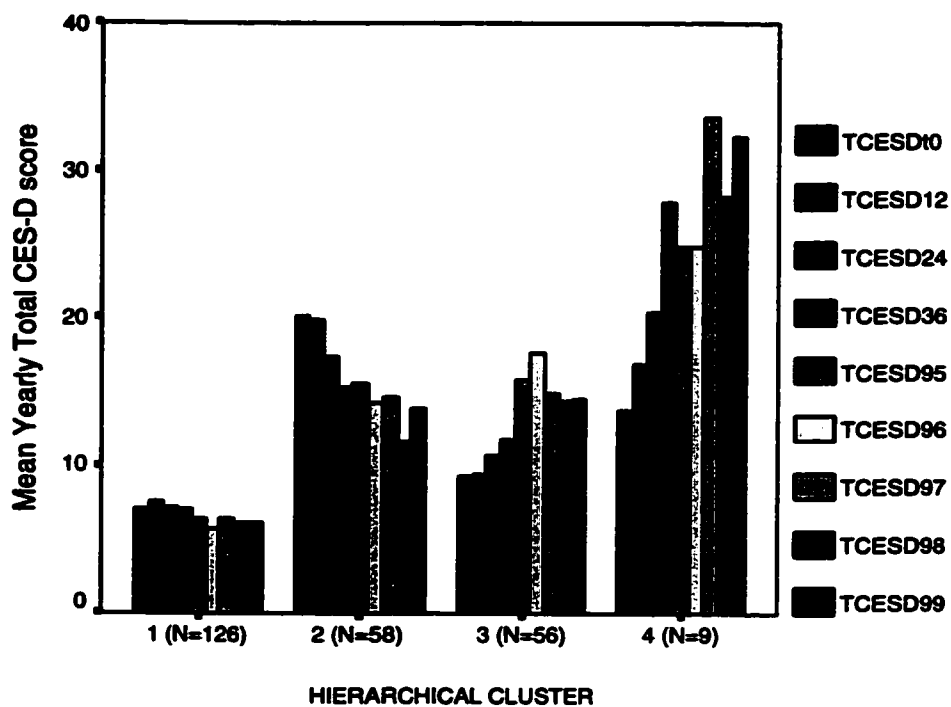


Figure 16  
Algorithm: Options for Parameters in Hierarchical Agglomerative Clustering



Ward's Method, Squared Euclidean Measure, RANGE99 as measure of variability, standardized Z-transformation of all variables

Figure 17  
Mean Yearly CES-D Score by Hierarchical Cluster

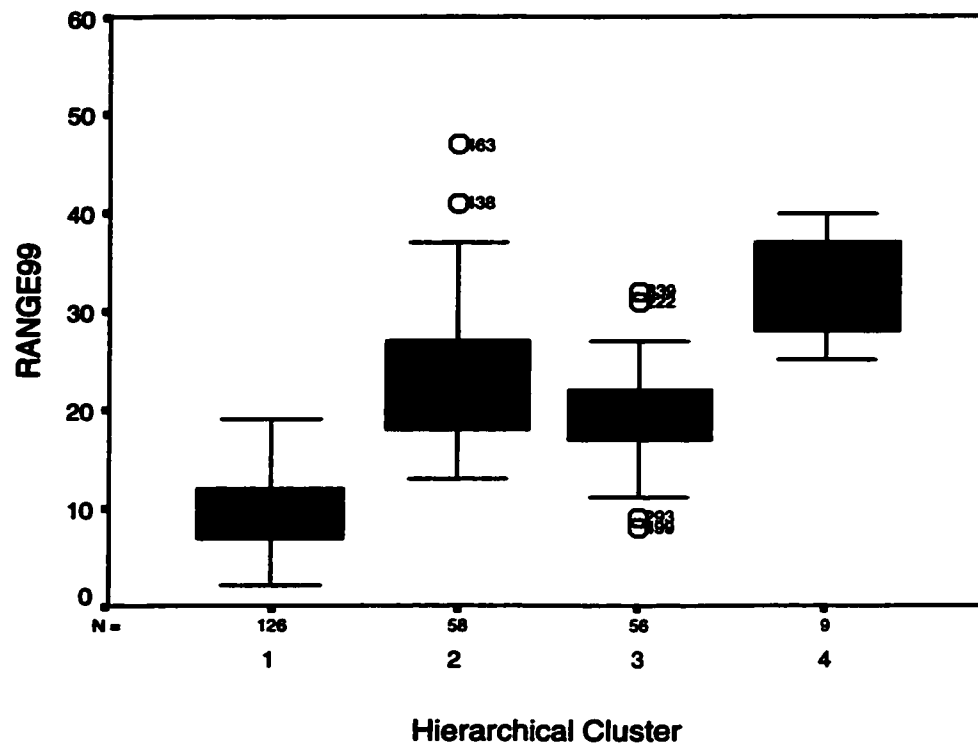


Figure 18  
Boxplot of RANGE99 by Hierarchical Cluster

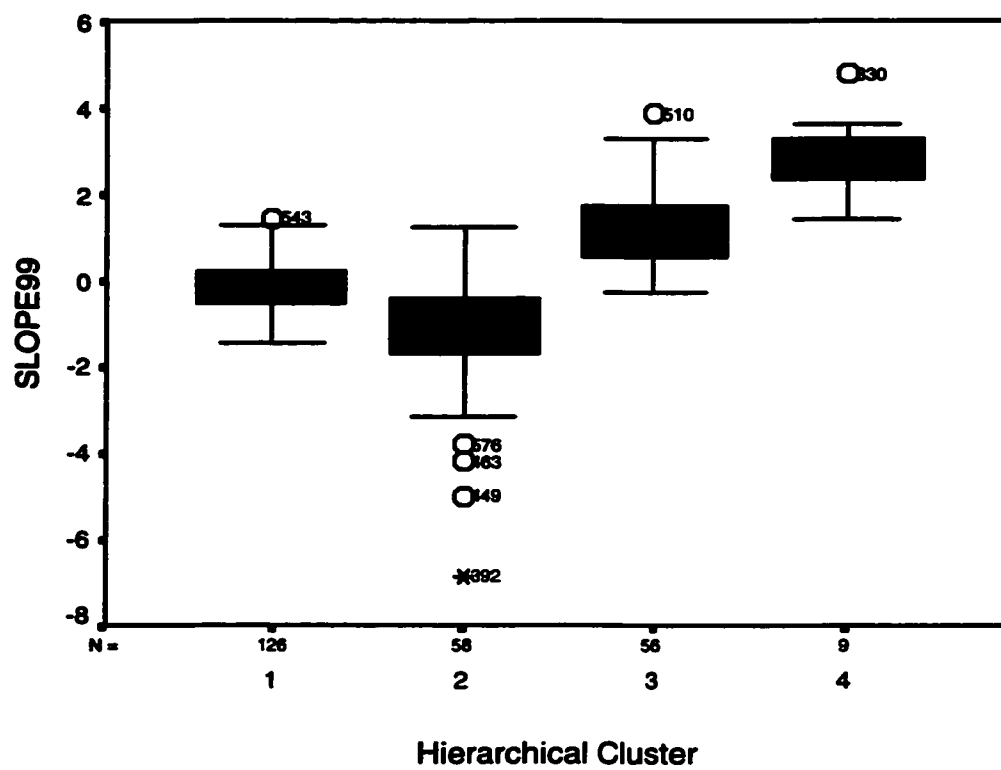
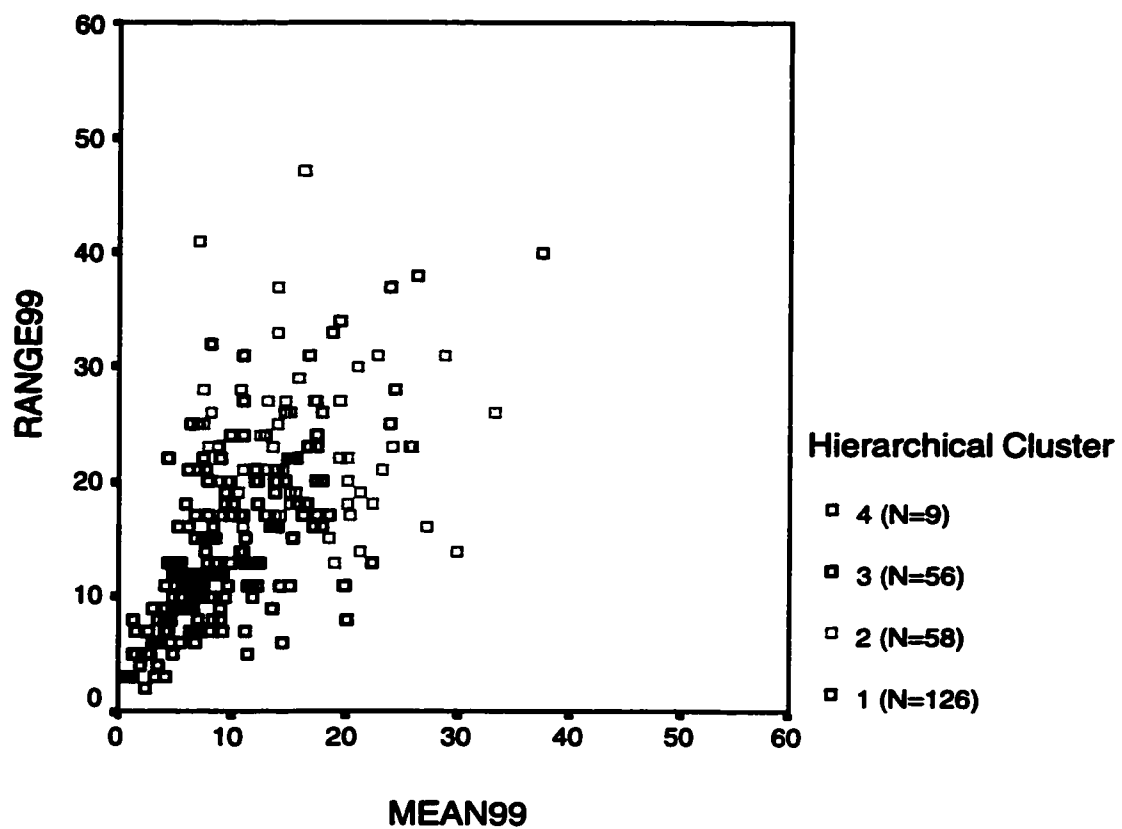
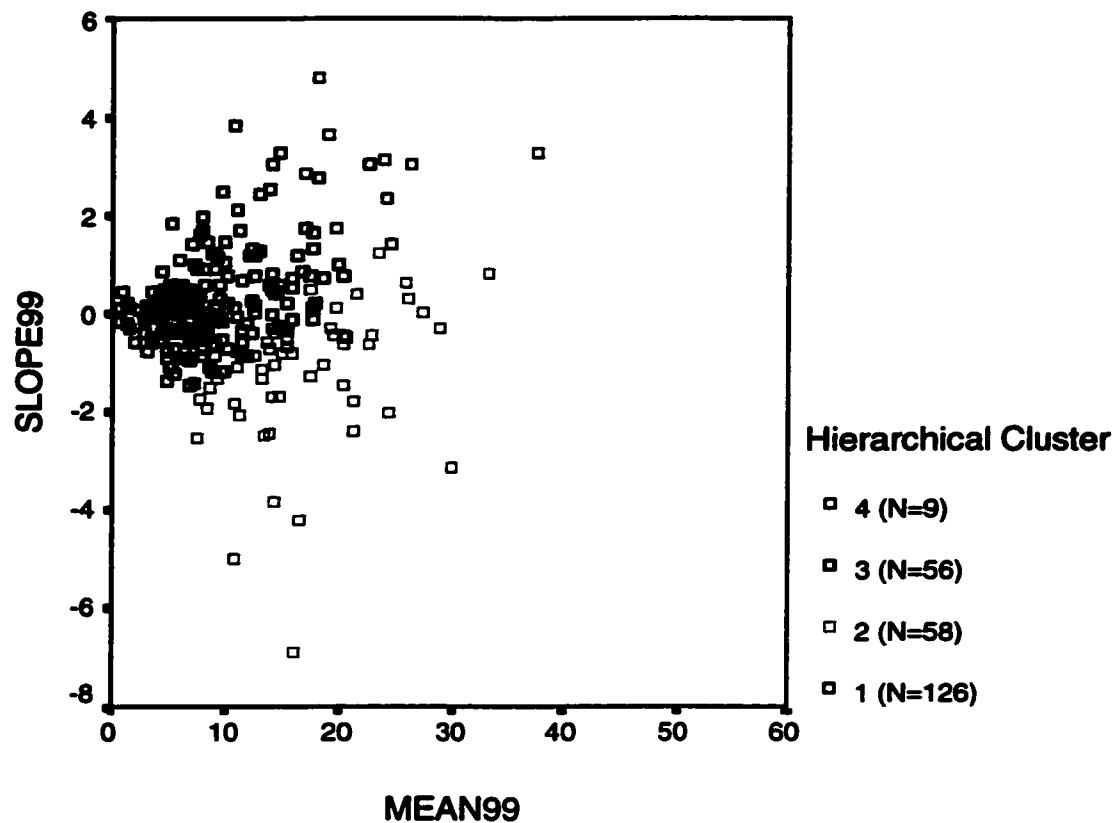


Figure 19  
Boxplot of SLOPE99 by Hierarchical Cluster



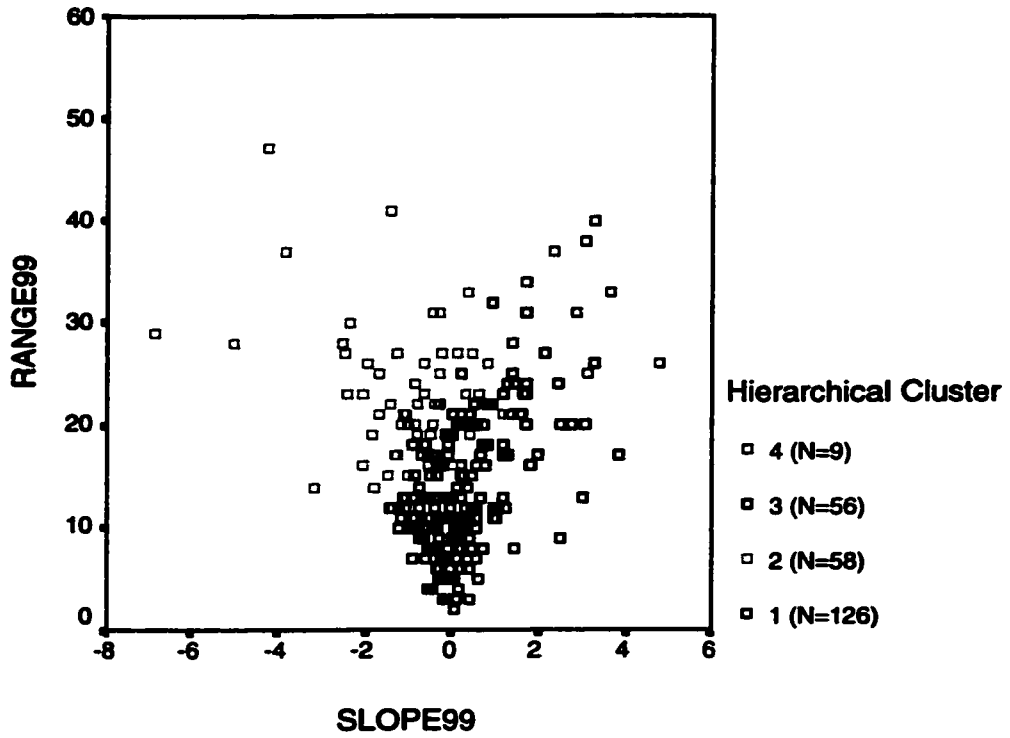
Ward's Method, Squared Euclidean Measure, RANGE99 as measure of variability,  
standardized Z-transformation of all variables

Figure 20  
Scatterplot of MEAN99 by RANGE99 by Hierarchical Cluster



Ward's Method, Squared Euclidean Measure, RANGE99 as measure of variability,  
standardized Z-transformation of all variables

Figure 21  
Scatterplot of MEAN99 by SLOPE99 by Hierarchical Cluster



Ward's Method, Squared Euclidean Measure, RANGE99 as measure of variability, standardized Z-transformation of all variables

Figure 22  
Scatterplot of SLOPE99 by RANGE99 by Hierarchical Cluster

**Table 5**  
**Iterative Or K-Cluster Final Cluster Centers:**  
**9-Cluster Solution; N=216**

	1 N=25 Stable MOD*	2 N=30 Stable LOW	3 N=31 FALL- ING LOW	4 N=2 RARE-1	5 N=2 RARE-2	6 N=22 RISE & FALL	7 N=78 Stable NON**	8 N=16 RISING LOW	9 N=10 FALL & RISE HIGH
TCESD T0***	14	7	20	45	11	9	6	7	26
TCESD 12	19	17	13	13	15	6	5	8	28
TCESD 24	17	9	14	16	39	8	7	10	24
TCESD 36	20	11	11	16	38	9	5	10	21
TCESD 95	26	7	11	34	15	18	5	10	19
TCESD 96	23	8	10	6	43	23	4	10	16
TCESD 97	20	9	11	7	48	14	5	7	31
TCESD 98	16	9	10	7	24	11	4	23	23
TCESD 99	21	7	11	5	35	10	5	14	27
SLOPE 99	.8115	-.3844	-.9619	-4.0071	2.7943	.8632	-.1241	1.0903	-.1442

\*Mod = moderate

\*\*Non=Non-depressed

\*\*\*Total CES-D score at entry to study at Time Zero

Table 6  
Summary Measures Of K-Cluster 9-Solution; N=216

CLUSTER		MEAN99	RANGE99	SDCESD99	CVCESD99	SLOPE99
<b>1</b> N=25	Mean	19.190087	20.72	7.429221	.3906615	.811469
<b>Stable Moderate</b>	Std. Deviation	2.7855112	5.79	2.0889298	.1074435	1.071726
<b>2</b> N=30	Mean	9.7054233	14.60	5.080558	.5312952	-.384412
<b>Stable Low</b>	Std. Deviation	2.1261080	5.09	1.7971846	.1860196	.664903
<b>3</b> N=31	Mean	12.653801	18.39	6.669514	.5458831	-.961943
<b>Falling Low</b>	Std. Deviation	2.2545705	6.31	2.5269237	.2375898	.811791
<b>4</b> N=2	Mean	15.3472222	42.00	13.9954393	.9088421	-4.007100
<b>Rare-1</b>	Std. Deviation	1.5517065	7.07	2.3446584	.0608841	.278317
<b>5</b> N=2	Mean	30.6944444	38.50	14.1506466	.4910787	2.794250
<b>Rare-2</b>	Std. Deviation	9.6245090	2.12	1.1592230	.1917485	.666731
<b>6</b> N=22	Mean	11.341495	19.55	6.655514	.6252761	.863200
<b>Rise &amp; Fall</b>	Std. Deviation	2.7855173	5.57	1.8387439	.2448691	.841035
<b>7</b> N=78	Mean	5.0702451	10.78	3.707947	.7799153	-.124059
<b>Stable Non-depressed</b>	Std. Deviation	1.9260337	6.43	2.1295614	.3630016	.671854
<b>8</b> N=16	Mean	10.417782	17.75	6.144917	.6019712	1.090342
<b>Rising Low</b>	Std. Deviation	2.6682300	6.56	2.3428682	.2095238	.864907
<b>9</b> N=10	Mean	24.675794	24.20	8.979862	.3691910	-.144176
<b>Fall &amp; Rise High</b>	Std. Deviation	4.1745153	7.05	3.3893455	.1390717	1.622924
<b>Total</b> N=216	Mean	10.7115767	16.13	5.6710783	.6173172	8.45340E-03
	Std. Deviation	6.2897194	8.07	2.9244367	.3034829	1.139835

Table 7  
Summary Measures Of Hierarchical Cluster 4-Solution; N=249

Cluster #	MEAN99	SDCESD99	CVCESD99	RANGE99	SLOPE99
1	6.7488042	3.5459183	.6162691	10.00	-.169021
STABLE NON*	N 126	126	126	126	126
	Std. Deviation	1.2737783	.3011145	3.73	.565828
2	16.4733688	8.2711671	.5791745	22.79	-1.187772
FALLING MOD**	N 58	58	58	58	58
	Std. Deviation	2.5111564	.3186120	6.61	1.405016
3	12.9666561	6.9197053	.6170438	19.64	1.168580
RISING LOW	N 56	56	56	56	56
	Std. Deviation	1.5847335	.3044158	4.63	.931486
4	23.2403440	12.3459703	.5539848	32.44	2.909222
HIGH	N 9	9	9	9	9
	Std. Deviation	2.5369278	.1411234	5.36	1.007106

Ward's method, squared Euclidean measure, RANGE99 as measure of variability,  
standardized z-transformation of all submitted variables

\*Non-depressive symptom level

\*\* Moderate depressive symptom level

**Table 8**  
**Descriptive Statistics Of Total Ces-D Score**  
**By Menopausal Transition Stage 1999; N=190**

	Pre*	Early*	Mid*	Late*	PM*	HRT*	OC*	Other*
Mean	--	11.41	9.38	13.49	7.67	9.14	11.28	13.42
SD	--	10.21	8.18	9.68	7.21	8.21	10.84	6.61
N	0	43	53	21	9	35	18	11

No significant differences between Menopausal Transition Stage 1999 (MTS99) by ANOVA;  
 F=1.087 (df=6,183) p=.372

\*MTS99 coding

Pre = Pretransition

Early = Early transition

Mid = Middle transition

Late = Late transition

PM = Postmenopause

HRT = HRT use, including known postmenopausal and unknown MTS stage women

OC = OC use

Other = all other categories

Table 9  
Yearly Anova Of Total Ces-D Scores By Menopausal Transition Stage \*

Statistics	TCESD 10	TCESD 12	TCESD 24	TCESD 36	TCESD 95	TCESD 96	TCESD 97	TCESD 98	TCESD 99
N	205	202	204	184	109	169	187	181	190
F-statistic	.558	.938	1.134	1.271	1.227	.370	1.006	1.611	1.087
P value	.811	.493	.335	.240	.269	.977	.449	.081	.372
Degrees of freedom	8, 196	9, 122	12, 191	12, 171	14, 194	13, 155	14, 172	14, 166	6, 183

\*Menopausal Transition Stages

Pre = Pretransition

Early = Early transition

Mid = Middle transition

Late = Late transition

PM = Postmenopause

HRT = HRT use, including known postmenopausal and unknown MTS stage women

OC = OC use

Other = all other categories

**Table 10**  
**Yearly Comparison Of Total CES-D Scores Between Late Transition**  
**And All Other Menopausal Transition Stages Combined**

Statistics	TCESD 10*	TCESD 12	TCESD 24	TCESD 36	TCESD 45	TCESD 54	TCESD 63	TCESD 72	TCESD 81	TCESD 90	TCESD 99
N late	1	6	10	13	9	16	22	24	21	21	21
N other	215	207	205	182	111	164	167	157	169	169	169
t-statistic	.193	.760	-.400	1.868	-.028	.340	1.558	-.705	-1.570	-1.570	-1.570
P value	.847	.448	.689	.063	.978	.734	.121	.482	.118	.118	.118
Degrees of freedom	214	211	213	193	118	178	187	179	188	188	188

\*Total CES-D scores at time zero, entry to study

**Table 11**  
**History Of Sexual Abuse/Assault By K-Cluster 9-Solution;**  
**N=174**

K-Cluster	Negative History of Abuse	Positive History of Abuse	Total
1 Stable MOD Adjusted Residual	12 (63.2%) -.7	7 (36.8%) .7	19
2 Stable LOW Adjusted Residual	15 (65.2%) -.5	8 (34.8%) .5	23
3 FALLING LOW Adjusted Residual	15 (62.5%) -.9	9 (37.5%) .9	24
4 RARE Adjusted Residual	1 (50.0%) (omitted)	1 (50%) (omitted)	2
5 RARE Adjusted Residual	2 (100.0%) (omitted)	0 (0%) (omitted)	2
6 RISE & FALL Adjusted Residual	14 (70.0%) 0	6 (30.0%) 0	20
7 STABLE NON Adjusted Residual	50 (76.9%) 1.5	15 (23.1%) -1.5	65
8 RISING LOW Adjusted Residual	11 (100.0%) 2.2	0 (0%) -2.2	11
9 FALL & RISE HIGH Adjusted Residual	2 (25.0%) -2.8	6 (75.0%) 2.8	8
<b>TOTAL</b>	<b>122 (70.1%)</b>	<b>52 (29.9%)</b>	<b>174</b>

Pearson  $\chi^2$  (excluding two rare clusters #4 and #5) = 15.23 (df 6), p=.019

**Table 12**  
**Historic Trend In**  
**Percent Using Prescriptive Antidepressant Drugs**

<b>CES-D year</b>	<b>Total % of sample Anti-depressant use</b>
<b>Entry to study (t0)</b>	<b>4.5</b>
<b>12 months</b>	<b>3.0</b>
<b>24 months</b>	<b>4.5</b>
<b>1995</b>	<b>8.0</b>
<b>1996</b>	<b>8.5</b>
<b>1997</b>	<b>12.0</b>
<b>1998</b>	<b>13.0</b>
<b>1999</b>	<b>19.4</b>

## **CHAPTER 5: PHENOMENOLOGICAL ANALYSIS**

**This chapter reported research aims, specific design, methods, results, and focused discussion for a phenomenological analysis, including a review of the analytic method in case this method were not familiar to the reader. General information about the parent study and general methods for the current study have been provided in Chapter 3. A general discussion has been provided in Chapter 6.**

### **Aims**

- 1. The primary purpose of this study was to describe and characterize longitudinal patterns of depressive symptoms in midlife women before, during, and after the transition to menopause.**
- 2. The secondary purpose of this study was to explore relationships between longitudinal patterns of depressive symptoms and selected predictors of these patterns or particular changes within patterns. Antidepressant use was expected to be a confounder, and the specific predictors were MTS and life events or stressors. While the menopausal transition might be classified as a life event or stressor, it has been separated conceptually because of its classic role in the study of midlife women.**

### **Review Of Analytic Method**

**In order to achieve the exploratory and descriptive aims of the study, non-inferential methods were required. In addition, methods had to be designed or adaptable to description of longitudinal patterns of repeated measures over time. One of the difficulties with analyzing longitudinal data sets containing three or more time points of repeated measures is the complexity and/or obscurity of the statistical approaches**

available (Collins & Horn, 1991). Usually, hundreds if not thousands of data points are required as well (Gershenfeld & Weigend, 1997).

In contrast, phenomenological analytic methods are ideal for non-statistical pattern identification, or more accurately, pattern creation (Fischer, 1975; Lincoln & Guba, 1985, Chap. 2, 12). These methods also are ideal for theory development (Fischer, 1975; Lincoln & Guba, 1985, Chap. 12). Phenomenological analyses can be called naturalistic because they use a holistic approach to studying a phenomenon within its context (Fischer, 1975; Giorgi, 1975; Tinkle & Beaton, 1986). In brief, phenomenological analysis is a holistic creation of value-bound understanding where interactions between investigator and investigated are acknowledged, continuously developing theory is individualized in a described context, and all elements are mutually interactive without designation of temporal cause and effect (Fischer, 1975, p. 150; Giorgi, 1975; Lincoln & Guba, 1985, Chap. 1).

Important features are holism, meaning, and context. The whole pattern of repeated measures of the CES-D was analyzed. The expected confounder of antidepressant use and predictors MTS and life events and stressors were part of the whole phenomenon being analyzed, and important aspects of the context. Context also included, for each individual measurement, the other repeated measurements of CES-D as well as time passing, or linear time. In phenomenological research, the total context of the conduct of the research is also considered important and expected to be made explicit, including what the researcher might bring to the process.

Phenomenological analysis recognizes that results, in this case longitudinal patterns, are a product of the interaction between researcher and what is studied. “The goal is to describe the phenomenon not with instrumental or other final precision, but so that it becomes transparent—recognizable through its varied manifestations in various contexts” (Fischer, 1975, p. 149).

Phenomenological analysis offers several advantages: it is non-statistical, readily understandable, and capable of elucidating patterns that other approaches cannot. It allows for holistic and intuitive understanding of patterns. In the current study, phenomenological analysis allowed for recognition of patterns that simultaneously and holistically related all of the data points over time to each other. Most especially with the longitudinal data set of the current study, the phenomenological method was the only approach that allowed for identification of complex patterns, where a single type of pattern included changes over time within it.

Researchers unfamiliar with phenomenological methods might not recognize the scientific rigor of this approach (Fischer, 1975; Giorgi, 1975). In natural or positivist methods, the focus would be on determining reliable criteria and assigning individual patterns to categories, establishing inter-rater reliability, and counting the number of examples of each pattern (Kazdin, 1982, Chap. 10). In phenomenological research, trustworthiness and transferability can be seen as analogous to the positivist research concepts of validity and reliability. With phenomenological methods, trustworthiness is determined by the reader, based on the quality and completeness of the description of the context and meaning and credibility of the results (Fischer, 1975; Giorgi, 1975; Lincoln

& Guba, 1985, Chap. 11). Transferability is the burden of the researcher applying results from one phenomenological study to another (Giorgi, 1975; Lincoln & Guba, 1985, Chap. 11). The transferability of this phenomenological analysis to the cluster analyses of Chapter 4 is considered in the general discussion (see Chapter 6).

For researchers used to a logical-positivist or empiric-analytic approach, a phenomenological analysis might be unfamiliar or perhaps thought of as a preliminary one, providing the researcher only with a means of becoming familiar with the data. However, phenomenological analysis is a recognized and valuable approach in and of itself, often employed in the fields of psychology and nursing (Giorgi, 1975; Wismont, 2000). In this case, a phenomenological analysis of longitudinal patterns of depressive symptoms in midlife women was undertaken for its own value, as well as to complement the cluster analyses (see Chapter 4).

Lastly, some operational aspects of the current study's analysis might be similar to those of case studies (Yin, 1984) or single case research designs with multiple baselines (Kazdin, 1982, Chap. 6) in that extensive information about MTS and life events and stressors were considered in terms of the outcome variable repeatedly measured over time. Nevertheless, single case designs such as described by Yin and Kazdin are constructed within a positivist philosophy of science. That is, an objective researcher seeks to discover a valid and reliable external truth. Phenomenological analytic methods are constructed within a phenomenological philosophy of science, which has already been briefly reviewed.

## **Research Design**

The current study, conducted with data from the SMWHS, was exploratory and descriptive in design. Description was focused on the longitudinal pattern of repeated measures over time. The time interval was approximately annual and the number of time points was six or more, up to 10.

## **Methods**

### **Sampling Frame**

The sampling frame consisted of all CES-D scores from all participants in the SMWHS, the parent study.

### **Sampling Procedure**

Women with at least six CES-D scores through early 2000 (through CESD00) continuous enough to determine a pattern were eligible.

### **Tools And Measures**

Simple line graphs of each woman's set of scores were inspected. Each Y axis (ordinate) was fixed from 0 to 60 to plot CES-D scores. Each X axis (abscissa) plotted time in number of years. Time was calculated by using the date written on the annual questionnaire or its postmark. The year of entry to the study was set to zero; each subsequent date of CES-D completion was calculated in years since entry to the study.

Summary variables of each individual woman's longitudinal set of repeated CES-D scores were calculated. These variables included a woman's mean of all her CES-D scores through 2000, labeled MEAN00; the range of all her CES-D scores through 2000, labeled RANGE00; and the slope of a simple fitted line, labeled SLOPE00, regressed

through all her CES-D scores through 2000 graphed along a weighted time axis. These summary measures or computed variables have been referred to in all capital letters with the year-designation from the final yearly CES-D score used in the computation at the end of the name. Summary variables allowed a representation of longitudinal pattern in the following ways: MEAN00 was a reflection of general level of scores; RANGE00 was a representation of the overall variability or lability of depressive symptoms. The overall sense of trend over time was summarized as SLOPE00, positive if the scores were generally worsening and negative if the scores were generally improving.

A compilation of life events and stressors that the women provided in a free form was conducted, expecting that associations with or even causes of changes in CES-D scores could be identified. The information about life events and stressors was obtained from otherwise unsolicited handwritten comments on the yearly questionnaires and the reverse side of calendars (see Appendix A). The instructions on the back of the calendars read:

**Record any unusual events including surgery (named), illnesses and medications, started or stopped. Include major events regarding health, work, residence, relationships, financial, legal matters, parenting. If you start, stop or change hormones please record this with the date.**

A second request on the reverse side of the calendars is to record address or phone changes, which would record moving as a life event. The majority of calendars do not have any recording on the back. However, some women used the opportunity to record so many life events that they needed extra space. It was expected that the importance and salience to the individual woman would be particularly great for life events and stressors that she bothered to write in as marginal notes on a questionnaire of otherwise structured

items and on the blank back of a calendar throughout the year, and therefore particularly appropriate for phenomenological analysis. Menstrual cycle calendars also included information about MTS and use of HRT. Use of antidepressants was recorded in diaries every year and in annual questionnaires starting in 1998 (see Appendix A). Responses to open-ended interview questions obtained at entry to both the first and second NINR grants were also included if pertinent to life events and stressors and MTS. A narrative of life events and stressors, MTS, and antidepressant use has been provided for women whose graphs are presented as examples of a particular pattern.

#### Analytic Procedures

The analysis was started as a method of becoming familiar with the data, at an early date before cluster analysis was known to the author, and when only CES-D scores through 1998 were available. Some of what became the most common pattern labels were identified early (stable near zero, Decreasing Oscillation Falling, Decreasing Oscillation No Trend, and Increasing Oscillation Rising—see Results later). A return to the phenomenological analysis occurred after the series of cluster analyses were performed (see Chapter 4). Results from the cluster analyses would be expected to influence phenomenological results. In particular, the specific levels of non-depressed, low, moderate, and high from the cluster analyses were likely to have been influential.

The determination of pattern was conducted as a visual gestalt. Different programs for line graphs were used, some leaving missing segments if a time point had a missing CES-D score and some completing the connecting line between every consecutive two points regardless of missing intervening data points. Some programs

fixed the distance between data points as constant (SPSS Sequence), and one (SPSS Curve Estimation) adjusted the distance for the specific value of time in years. The former did not include the regression line. For visual pattern analysis of line graphs the CES-D was considered a continuous variable.

All the graphs were reviewed by the investigator in a single session. A record was kept of the development of pattern labels and the pattern assignment of each graph. Three such sessions were required for refinement of pattern identification and labeling. During the third session very few additions or changes were made to the pattern labels and categories. Thus, it was felt that additional review sessions would not provide further pattern development. After the third session the assignments of graphs was compared to the assignments during the second session. There was modest consistency in pattern assignment.

After discerning each woman's pattern, reference lines at 16 and 24 were added to graphs in order to indicate the cut-offs for major depression according to Radloff (1977) and Roberts et al. (1991), the lowest and highest cut-offs, respectively, in literature reviewed up to the date of this analysis. A simple regressed line was also fitted to the data set for each woman. While reference lines and regression lines were occasionally useful, they were not generally used in the procedures which discerned the woman's pattern. They have been included in the graphs shown here to assist the reader.

## **Results**

### **Sample**

There were 223 women with at least six CES-D scores through CESD00, the latter being the CES-D score obtained in early 2000. However 18 of these women did not have CES-D scores measured repetitively and continuously enough so that a pattern could be discerned; these women were missing at least three consecutive years of scores in the middle of the study (see Table 4). The final sample, therefore, consisted of 205 women.

### **Longitudinal Patterns**

Three dimensions of patterns were identified. First was the shape; there were waveform, non-waveform, or combination patterns. Second was the trend; there were flat, rising or falling patterns, reflecting SLOPE00. Third was the level; there were non-depressed, low, moderate, or high patterns, reflecting MEAN00. In addition, waveform patterns were characterized as Seesaw, Slow Oscillation, or Rise & Fall (or Fall & Rise), Increasing Oscillation, or Decreasing Oscillation. The terms seesaw and oscillation were considered synonymous. Non-waveform patterns included phenomena such as Stairstep Up (or down), single Blips (up or down), or double Blips, labeled as occasional good (or bad) year. Combinations were those that contained at least two of these patterns; that is, combinations were not readily classifiable as a pattern.

### **Waveform Patterns**

In reviewing graphs, often the sense of an oscillation or waveform predominated. Waveforms have several characteristics, including amplitude, period, frequency, and

wavelength (Giancoli, 1998, p. 310, 325). While amplitude is defined as the height or depth of the wave measured from a central point of equilibrium (p. 325), for present purposes the term has been applied to the entire distance from peak to trough, or technically twice the amplitude. Period was defined as the time required for one complete cycle (1988, p. 310) and visually has been used here to refer to rapidity of the waveform. Frequency refers to the number of complete cycles per unit of time (1988, p. 325). Wavelength refers to the distance from beginning to end of one complete waveform. Since the horizontal or x-axis determinant of the waveforms in the current study is a fixed ten-year time frame, frequency and wavelength can be synonymous here. For the current study waveforms were placed within a fixed range of CES-D scores from 0 to 60 on the y-axis, and therefore vertical placement or level was also an important characteristic. These characteristics of the waveform patterns were used to describe and distinguish between them.

The first characteristic was a visual sense of RANGE00, the amplitude of the waveform. Some patterns had a small mountain or seesaw pattern; still others had a large mountain or seesaw pattern. The former was labeled Minor Seesaw, the latter Major Seesaw. Differences between consecutive measures were studied from a print out of all scores for each woman after the visual gestalt to determine pattern. Consistently there were 9-15 points difference at least once in the 10 years between two consecutive years for Minor Seesaw. Major Seesaw patterns had more than 15 points absolute difference between two consecutive years at least once. Even though it was not a waveform pattern,

stable flat could be seen as part of a continuum of decreasing variability (amplitude) within the seesaw patterns, from Major Seesaw to Minor Seesaw to stable flat.

A second characteristic was period, operationalized as the number of waves in the 10-year time frame along the X-axis. However, if amplitude, the first characteristic, were zero or flat this second characteristic of period was obviated. Period could be seen as theoretically lengthening progressively from Seesaw to Slow Oscillation to Rise & Fall (or Fall & Rise) patterns, within each of the groupings of amplitude.

Period reflected how quickly the waveform changed direction, and therefore implied wavelength and frequency. Seesaw patterns showed a change in direction nearly every year, usually with no more than two years in a row continuing in the same direction. The wavelength was two years in that case. To put it another way, a seesaw pattern showed a rise to and drop from each peak lasting only one or occasionally two years. Thus seesaw pattern had at least three mountains over the 6-10 years. Slow Oscillation was a longer period. Two or three mountains over the 10 years gave a sense of slow oscillation, especially if the rise to or drop from a peak occurred over at least two or more years. The observance of only one mountain (or valley) in the six to 10 years was described as Rise & Fall (or Fall & Rise).

Level, a third characteristic that was used to determine pattern, had two dimensions, level and trend. Level was generally divided into four subcategories by visual inspection: Non-depressive or non, low, moderate, and high. The non-depressive level usually had all CES-D scores lower than 9. Low usually had CES-D scores, or at least a MEAN00, around 9 to 12. The moderate level was around 15 to 17. High was at

a level above 20 for the most part. While both dimensions of level were determined phenomenologically by visual gestalt, the patterns created by the author could be seen to reflect the 50<sup>th</sup> percentile for the variable MEAN00 at 9.0, the 75<sup>th</sup> percentile at 14.0. The mean value was 10.6 and standard deviation was 6.3. RANGE00 had a mean of 16.5, standard deviation of 8.1, 25<sup>th</sup> percentile at 10.0, 50<sup>th</sup> at 16.0, and 75<sup>th</sup> at 21.0. All descriptive statistics were calculated from the sample of 223 women with six or more annual CES-D scores. The second dimension of level was trend or slope, again determined by visual inspection. If the level were basically unchanging this characteristic was seen as no trend; that is, the waveforms appeared to oscillate around a stable mean value. If the overall level changed it was seen as either rising or falling.

Many combinations could be made from each of these three characteristics. Theoretically, all combinations of all waveform characteristics would yield 2 amplitudes X 3 frequencies X 4 levels X 3 trends for 72 waveform patterns. However, not all combinations are actually possible. For example, there could not be waveform patterns at the non-depressed level. This is because if there is enough change to be classified as seesaw, the high points raise the mean level above non-depressed. There could also not be minor or Major Seesaw or Slow Oscillation rising at the non or low levels, because if they were rising the MEAN00 would no longer be non or even low.

Individual patterns have been presented here within trend by frequency and amplitude. While in general there were examples of each subpattern at each level, for the sake of conciseness only one example of each general pattern has been provided. A narrative combining information of MTS, life events and stressors, and antidepressant use

had been included for each example. A matching figure has been included as an example of each major category as well. In the figures a line connects CES-D total scores from one measurement to the next starting with entry to SMWHS at time zero. The straight line is a fitted simple regression through the data points. There are two reference lines, one at 16, the standard cut-off for the lower limit indicative of major depression, and one at 24, the higher cut-off indicative of major depression (see Chapter 3). For each woman the y-axis (total CES-D score) has been fixed to scale from 0 to 60 and the x-axis (time in years) has been fixed to scale from 0 to 10. MTS coded by letter have been added (U=unknown, P=pretransition, E= early transition, M=middle transition, L=late transition, PM=postmenopause, HRT or H=hormone replacement use, OC=oral contraceptive use, and O=other such as hysterectomy, radiation or chemotherapy). Occasionally, in order to assist the reader, the corresponding year that the CES-D score was measured has been added beneath the code letter.

**No trend.** Waveforms without trend had six categories, all of which demonstrated a SLOPE00 of approximately zero: Minor Seesaw, Major Seesaw, Slow Oscillation, Rise & Fall, Increasing Oscillation, and Decreasing Oscillation.

**Minor Seesaw:** The example in Figure 23 demonstrated Minor Seesaw at a Low level No Trend, a fairly common pattern. There were very few women demonstrating Minor Seesaw at a Moderate level No Trend, and none at the High level.

**Life events & stressors:** The woman demonstrating a Minor Seesaw pattern in Figure 23 started HRT between years 2 and 3, and between years 8 and 9 started a successful weight loss program still in effect at year 10. She also experienced

advancements in her work between years 0 and 1 and again between years 2 and 3, family losses and a car accident between years 4 and 5, major travel between years 7 and 8, accidents between years 8 and 9 and again between years 9 and 10; the latter time span also included moving twice. There was no apparent relationship between these events and the years of rise as opposed to the years of fall. MTS also seemed unrelated to CES-D score or change in score.

**Major Seesaw:** There were two levels where a Major Seesaw pattern was demonstrated in this sample, low and high, although Major Seesaw Moderate No Trend is theoretically possible. An example of Major Seesaw at a Low level has been shown in Figure 24.

**Life events & stressors:** This woman experienced the death of her father and unexpectedly of a friend in the months before year 2. Separate crises with her mother's health occurred before low scores at years 3, 8, and the high score at year 7. Yearly life stressors occurred for this woman concerning her work, such as being laid off, traveling for 19 months at a time, and quitting. Changes were mentioned before year 2 (a rise) and year 3 (a fall) in terms of dieting and weight loss. This woman had repeated problems with migraines after year 2 and a minor motor vehicle accident with facial trauma five months before the rise of year 9. Before the year 5 fall in score, she discontinued the use of an antidepressant drug that she had been taking for her migraines. MTS seemed unrelated to CES-D score or change in score.

A remarkable aspect of these observations is the lack of correspondence between life events and annual CES-D score. While it might make sense that a rise in her final

CES-D score at year 9 came just after she fell in love with another person and her life partner moved out, the “most painful experience of my life,” this rise did not reach the level of years 2 or 7. Such a CES-D score would not even have been categorized as depressed using a cut-off value of 16.

**Slow Oscillation:** Slow Oscillation was observed as having two complete or nearly complete waveforms in the approximately ten years. Again, this pattern occurred at more than one level. An example of Slow Oscillation No Trend at a High level is presented in Figure 25.

**Life events & stressors:** This woman provided many details of her life most years, including even colds and flu. Just after entry to the study she started a new relationship, which she noted was stressful between year 1 and 2, and with which she continued to note problems especially during years 3-5. The relationship ended between years 5 and 6 when her scores were somewhat stable, although they continued to live together through year 9. She recorded such health events as an HSV (herpes) episode year 3-4, hospitalization for vertigo year 4, gastrointestinal problems year 5, and subsequent diagnosis with IBS (irritable bowel syndrome). She started Xanax for anxiety year 6, tried acupuncture for anxiety the next year, and between years 6 and 7 added Paxil and Klonopin with a subsequent marked decrease in her score for the second time. If earlier scores were not available, the second fall might be attributed to the antidepressant drugs. By the final score she was again taking Xanax only. Between year 4-5 she started weekly psychotherapy. At her peak score year 5-6 she had to stop regular exercise because of health problems. The unchanging MTS was unrelated to CES-D score or

change in score. While this woman seems to have had many physical and depressive symptoms, there does not seem to be much explanation for the low scores at years 3 and 7-8.

**Rise & Fall:** Rise & Fall (or Fall & Rise) had only one complete waveform in the 10 years and could be seen as demonstrating even longer period than Slow Oscillation. However, the Rise & Fall designation required stability of up to two years before and after the waveform. Only one woman was characterized as Fall & Rise, a mirror of Rise & Fall, and has been omitted for brevity. An example of Rise & Fall has been shown in Figure 26, although because there was no CES-D score between years 3 and 5, the slow rise might have been an artifact.

**Life events & stressors:** During the stable time between entry to the study and year 1 this woman moved, traveled internationally twice, and experienced a parent's death. Between year 5 and 7-8 she annually recorded that she attributed depression, loneliness, sadness, oversensitivity and anger to menopause [transition]. Immediately before the high score at year 5-6 she had a 2-week fight with her husband. Between years 7-8 she reported work problems, becoming a union steward and quitting it. At the stable low years 8-10 she experienced two deaths, bronchitis, international travel, a cervical polyp, and her mother's stroke and requirement for complete bed-ridden care in this woman's own home. There was no apparent relationship between any of these stressors and the changes in scores. Despite this woman's own attribution, MTS also seemed unrelated to CES-D score or change in score.

**Increasing Oscillation:** Increasing and Decreasing Oscillation referred to the change over time in amplitude; that is, to increasing or decreasing seesaw-ness over time, respectively. Technically these could be categorized as combination patterns, flat becoming seesaw or seesaw becoming flat, or perhaps Major Seesaw becoming Minor Seesaw or vice versa. However, Decreasing and especially Increasing Oscillation were so common that they warranted recognition as distinct categories, occurring at low, moderate, and high levels. There were also patterns of Increasing and Decreasing Oscillation where the increase or decrease was not so much in amplitude as in frequency. Some women demonstrated these changes in oscillation in both characteristics simultaneously, that is, both decreasing amplitude and decreasing frequency (oscillation slowing) or increasing amplitude and increasing frequency (oscillation speeding up).

**Increasing Oscillation No Trend** was the largest subcategory in No Trend, and along with **Decreasing Oscillation Falling** one of the most common subcategories for all the women in this analysis. The pattern in Figure 27 was an example of **Increasing Oscillation at a Low level No Trend**.

**Life events & stressors:** For background information this woman was a self-employed artist and vegetarian who practiced regular meditation and exercise. At year 1 she had new work. At year 5 she answered an interview question that her health concerns included gaining weight. She also responded to an interview question at year 5 about “How well can you predict the start of each menstrual period?” with the additional information that she experienced behavior change, uncontrollable anger, and sleeplessness premenstrually. At year 6 she started taking phytoestrogens for symptoms

of sweating and sleep problems. At year 8 she reported a decrease in income, in spite of the lower CES-D score. There was no relationship between the scant information she provided and the increasing instability of her scores after year 3. MTS seemed unrelated to CES-D score or change in score.

**Decreasing Oscillation:** Decreasing Oscillation No Trend was also a fairly common pattern, occurring also at low, moderate, and high levels. Figure 28 showed an example of Decreasing Oscillation No Trend at a High level. While it might appear from the regression line to have a slightly positive slope, the difference between the intercept of the line and the final score is only six points (20-14). The regression line was not present in graphs during phenomenological analysis and no general trend was perceived at the time of reviews.

**Life events & stressors:** This woman took almost no medications or vitamins throughout the nine years of measurement. At year 1 she had started her own home business. Between years 1 and 3 her mother and father individually had several problems with health and eventually died requiring her to travel out of town for months. Yet she reported at highest score year 2 that she had improved her diet and exercise. In contrast at the low-symptom point year 4 she complained of constant back and neck pain. As her scores became more stable she reported a suicide of her son's friend and improvement in business. According to her own comment at year 7 she has "some periods of profound depression without any clear cause." MTS seemed unrelated to CES-D score or change in score.

**Rising. Minor and Major Seesaw, Slow Oscillation, Increasing and Decreasing Oscillation** were also patterns observed rising. Non, low, moderate and high levels in the rising and falling patterns were occasionally difficult to describe as they depended as much on SLOPE00 as well as MEAN00 for the individual woman. If the SLOPE00 had an absolute value of three or higher, there was so much change that all levels were exemplified at some point in time. If the SLOPE00 value was rising but less steep, then a general sense of the level could be determined. In general, women with an overall rising trend were few, in almost all of the subcategories. The exception was Increasing Oscillation Rising, the largest subcategory in Rising and one of the more common patterns overall.

**Minor, Major Seesaw; Slow, Decreasing Oscillation:** While there was only one example of Minor Seesaw Rising, in this case at a high level, theoretically this pattern could occur at various levels. There were a few more examples of Major Seesaw Rising, with steeper slopes than Minor Seesaw, as well as with the greater changes from year to year. There were also a few examples of Slow Oscillation Rising. Decreasing Oscillation Rising also occurred rarely, sometimes with decreasing oscillation of both amplitude and frequency. For brevity the rare rising patterns have not been shown.

The largest group within the rising category was Increasing Oscillation Rising. Women with this pattern showed, in effect, worsening symptoms with destabilization. Again there were women at different levels and with different steepnesses of SLOPE00. The example in Figure 29 demonstrated both increasing frequency and increasing level.

**Life events & stressors:** This woman reported she resumed full-time work between entry to the study and year 1, switched from aerobics to yoga between year 1 and 2, and finished a dissertation by year 3. Between year 3 and 4 she married her long-time partner and moved the next year. At year 5 she was looking for a job. In spite of all these events she was fairly stable. She did report increasing quantities of life events and stressors as instability and symptom-level increased. Her mother had health problems, and she had just been mugged at year 7. Separation and divorce occurred during the last three years of measurement. At year 9, the same score as year 3, she reported use of St. John's wort, and continued its use through year 10. Both her parents had increasing health problems at year 10 and she had mononucleosis. She had moved twice in the previous year.

The increasing instability and level might be attributed to the end of her marriage that was accompanied by worsening parents' health and even post trauma after being mugged. In spite of the events preceding the increasing instability and rise, this woman might exemplify the effects of at least certain life events and stressors on depressive symptoms. However, it cannot be determined whether this might be due to coincidence. Given the lack of change in early menopausal transition during the increasing oscillation, MTS seemed particularly unrelated to CES-D score or change in score.

**Falling.** The falling waveform category repeated the same subcategories as rising: Minor and Major Seesaw, Slow, Increasing and Decreasing Oscillation.

Minor, Major Seesaw; Slow, Increasing, Decreasing Oscillation: Seesaw patterns, either Minor or Major, with a Falling trend were quite rare. No Minor Seesaw

patterns were exemplified at the low level and few Minor Seesaw patterns were at the Moderate level. The remaining few women were Major Seesaw at the Moderate and High level. Slow Oscillation Falling was also a rare category. Increasing Oscillation Falling occurred at two levels, moderate and high, and was also rare. A woman's scores demonstrating such a pattern became less stable, even though her symptoms improved over the years. Examples of the rare Falling patterns have been omitted for brevity.

In mirror fashion to Increasing Oscillation Rising, the single largest Falling subcategory was Decreasing Oscillation. In fact, Decreasing Oscillation Falling was one of the most common patterns overall, along with Increasing Oscillation No Trend. For these women falling CES-D scores, or improvement in symptoms, were expressed with increasing stability. This pattern also occurred at various levels. The example in Figure 30 at the low level was representative.

**Life events & stressors:** This woman's background included a history of "severe PMS." She began a new writing career and new relationship between entry to the study and year 1, when score was rising. She experienced increased job responsibility and hours between years 2 and 3, also rising. At rising year 3 she had stopped school, gained weight, and "my son was shot three times." She moved between years 3 and 4, again between years 5 and 6, and again between years 8 and 9, with falling, rising, and stable scores, respectively. At year 5 she had started another new relationship, gotten her degree, and "overcame" a job situation of sexual harassment that was now better. At year 6, even though this was not a peak year, her 18-year-old daughter in college was pregnant and she wrote of having to "deal with this pain," and one of her sons went to prison. At

year 7 she was working on a master's degree. Between years 8 and 9 she traveled internationally and changed jobs. The reported intensity of events and their level of stress seemed to increase, especially year 6, yet she was becoming more stable. OC use might have been related to the decreasing oscillation in CES-D scores.

In the Falling subcategories some of the negative slopes were so large that it was difficult to label the pattern with a single level. In fact, the example of Decreasing Oscillation Falling in Figure 31 demonstrated the largest negative slope of the entire sample, -3.86.

**Life events & stressors:** This woman's background included frequent and disabling migraines occasionally requiring hospitalization. At entry to the study this woman was experiencing a "crisis," unable to do anything while waiting for a neighbor committed of sexually assaulting her sons to go to court. At her low-symptom point year 1 she had changed jobs. At her second peak year 3-4 the neighbor was in jail and her son was using drugs. Yet at her decreasing oscillation of years 4-5 she had many stressors, including her son's continuing drug problem, damage to home, thefts, plus a specific concern about her marriage. By year 6 she was postmenopausal and had pursued various therapies for her migraines. She complained of allergies at her lowest score year 7-8. It was possible that the sharp increase in CES-D scores between years 2-3 and 3-4 was due to HRT use, and that the sharp decline between years 3-4 and 4-5 were due to discontinuation of HRT and stabilizing as postmenopausal.

**Grand W.** A Grand W pattern was demonstrated by a few women. Theoretically there could also be a Grand M although no women were categorized as such. The

example of Grand W in Figure 32 differed from Slow Oscillation in that the center score though high was still much lower than the early and late high scores.

**Life events & stressors:** This woman's background included the death of one of her sons 18 months before entry to the study. During the first decrease between years 1 and 2 she moved, quit a job and got a new one. At the year 4 modest rise she was working 75 hours per week including a great deal of travel and diagnosed with depression, sleep apnea, and fibromyalgia. She started Ritalin at falling year 5 and during the next year stopped Prozac, started Redux, then restarted Prozac and switched to Paxil. At the low-symptom year 6 she had surgery that resolved "years of pain and problems." At the final high score she reported a death in the family, a diagnosis of cancer, closing her business and starting a new one, and a diagnosis of "Morphea," presumably a euphemism for drug addiction. MTS seemed unrelated to CES-D score or change in score, with the possible exception of two of the later low years while on OCs.

#### Non-Waveform Patterns

In addition to waveform, or mountain or seesaw patterns, there were also non-waveform patterns. These included stable and mostly flat at a particular level, and generally linear rising (or falling) patterns, blips and stairsteps.

Stable. Patterns that were classified as stable or flat if they generally showed less than 9 points difference between any two consecutive years across the entire set of total CES-D scores and no trend to rise or fall. While 9 points difference was determined phenomenologically, it was also noted that the standard deviation of RANGE00 was 8.1 and the 25<sup>th</sup> percentile for RANGE00 was 10.

Similar to waveform patterns, level was divided into four subcategories: Non-depressive or non, low, moderate, and high. After determination by visual gestalt, the non-depressive level was noted usually to have all CES-D scores lower than 9. Low usually had CES-D scores around 9 to 12. There was only one woman stable at a moderate level, with CES-D scores around 15 to 17. Stable High would have been CES-D scores stable at a level above 20 although there were no women in this sample with such a pattern. However, since phenomenological methods are used also to describe the possible range of patterns to assist theory development (Fischer, 1975, p. 148), a Stable High pattern is included theoretically in this non-waveform category. The stable category provided the single most common pattern, which was stable non-depressed. An example has been shown in Figure 33. No CES-D scores were missing for this woman; years 3, 4, 6, 8 and 9 were all total scores of zero.

**Life events & stressors:** The woman exemplifying Stable Non-depressed pattern identified many and complex life events and stressors. She struggled with cigarette and marijuana addiction (her own word) from entry to the study until between years 8 and 9, when she finally quit for good. Between years 6 and 7 she stopped smoking for the first time. During the next few years she restarted and stopped smoking several more times while also using herbs (St. John's Wort and Valerian root) and later antidepressants (Zyban switching to Paxil). Between years 3 and 4, when this graph became even more stable at the lowest possible total CES-D score of zero, this woman improved her job situation in several aspects (school completion, certification, new more responsible job, travel abroad to teach and work in her field), started HRT, and also switched from a

summer-time completion of the questionnaire to a winter one; the seasonal change would have been expected to have a worsening effect on the total CES-D score. This woman's mother became seriously ill between years 7 and 8. MTS seemed unrelated to CES-D score or change in score except that at the change from middle to late transition symptoms improved a negligible amount.

Two remarkable aspects of these observations are that 1) even though this woman did not evidence depressive symptoms she was prescribed antidepressants, although for smoking cessation not for depressive symptoms *per se*, and 2) she had many life stressors both positive and negative that had little effect on her CES-D scores at least at the times they were measured.

Flat or Stable Moderate was a category with only one exemplar from the SMWHS sample. Because the concept of stability at any level is demonstrated remarkably by this woman, a graph of her scores has also been included in Figure 34.

**Life events & stressors:** This woman was devoutly religious, with her pastor high up on her social support list, raising children who were pre-school aged at time of entry to the study. After year 4 she had no further menses. She provided the comment at year 5 that there were financial problems, coincident with a negligible rise in CES-D score nearly to its original value. MTS seemed unrelated to CES-D score or lack of change in score.

**Generally Linear Rising, Falling.** A few generally linear rising and fewer falling patterns were noted. Generally Linear Falling occurred at different levels, that is, starting at different levels and/or with different steepnesses of slopes. Generally Linear Rising

mirrored Generally Linear Falling. An example is omitted for brevity. Generally Linear Falling was the second largest group within the falling category.

An example of Generally Linear Rising is shown in Figure 35. However, even though ten years is a long time to follow women longitudinally, it might require a few more years before the full pattern of this woman's CES-D scores is demonstrated. If a general trend to rise continues, this graph would perhaps better be labeled as a Slow Oscillation Rising pattern. On the other hand, if the last three data points were not available, she would also be more clearly Generally Linear Rising. However, if in the next two or three years she continues to slowly fall, this might better be labeled as Rise & Fall.

**Life events & stressors:** This woman's background included a long-term diagnosis of hypertension. She married between entry and year 1 as well as finished her degree in education. At her lowest score at year 2 she had a new job with increased hours and responsibilities. Then at the 4-5 year point she explained the high number of work days missed by writing in that she was "tired, disillusioned with system." At the highest score, year 5-6, she had started HRT, had a diagnosis of angina, and complained of fatigue, stress, and money problems related to "going back to school" [before entry to study], and her husband's back problem and his subsequent disability. At slightly falling year 6 she started St. John's wort, progressing to cortisone treatments and Neurontin year 7, and Elavil and Valium between year 7 and 8. She was back to being a teaching assistant by year 7. At the final point she was not taking any antidepressants.

For this woman life events and stressors might be interpreted as causal or at least coincident with her rising scores, especially if stressors in the first few years are discounted. It is possible that the combination of work stressors, lack of support from her new husband plus increased stressors from his injury and disability produced the rising symptom score. Again, it is possible that the use of antidepressants alleviated the symptoms somewhat in the years after 5-6. MTS mostly seemed unrelated to CES-D score or change in score, except that the possible fall in the final three scores might be attributable to HRT use.

Blips. A Blip Up was a single year elevation of usually more than 9 points from at least one of the immediately preceding or succeeding years. The blip could occur early, late, or in the middle of the set of total CES-D scores. Because of the sampling in the parent study, the timing of the blip was somewhat arbitrary as to the woman's age or MTS. What distinguished an early Blip from Decreasing Oscillation and a late Blip from Increasing Oscillation was the isolated nature of the single year elevation. An example of a Blip pattern has been shown in Figure 36.

**Life events & stressors:** This woman's background included cholecystectomy and laminectomy several years before entry to the study, and a diagnosis of diabetes. She had four teen daughters at entry to the study. Her religious faith was prominent in the interview. The laminectomy was repeated between years 1 and 2. She had started Prozac, after breaking her leg between years 2 and 3, for a diagnosis of "low grade depression," in spite of apparently few depressive symptoms on her yearly questionnaires before then. She also started Provera between years 2 and 3. The only calendar on which

she included many events was for the year preceding the blip up when she received a diagnosis of bipolar disorder, started a new job, suffered a good friend's death, had a D&C followed by two courses of Provera, and had a daughter diagnosed with Bell's palsy. She also wrote at the blip up between years 5 and 6: "[I] want a cure to my depression." Between years 3 and 4 she lost work hours but was unable to get a better job until year 6. At the CES-D between years 6 and 7 she stopped Paxil and started St. John's wort for "menopausal depression" but switched to Nortriptyline before year 8.

Interpretation of this woman's annual CES-D scores is problematic in light of her diagnosis of bipolar disorder and prescriptive antidepressants. The information available to the parent study did not include evidence of mania or hypomania or a prescription for Lithium. Perhaps chance timing of CES-D administration or social desirability traits produced CES-D scores consistently in the non-depressed range. It is also possible that the quantity of life events and stressors overwhelmed the effects of antidepressant medication during the blip up year. MTS seemed particularly unrelated to CES-D score or change in score.

An occasional bad (or good) year could be seen as an extension of the Blip Up (or Blip Down), with the single year elevation (or drop) occurring twice in the six or more years, although far enough apart not to be identified as decreasing or Increasing Oscillation. Examples of occasional bad or good year are omitted for brevity.

Stairsteps. Stairstep Up (or down) was another pattern that warranted a separate category. The Stairstep label was applied to a pattern demonstrating dramatic and subsequently stable change in CES-D score from one year to the next, from a previously

stable flat pattern. The Stairstep label could be misleading, however. On the one hand there might be a true stairstep, reflecting a one-time change from one stable level to a different stable level in the woman's symptoms. However, it was also possible that the stairstep was an artifact of the combination of an extremely long period with a limited number of years of measurement; that is, the period was longer than the 6 to 10 years during which measurements were obtained. Half a mountain in 6-10 years could be described as a Stairstep, Up or Down. The graph in Figure 37 has provided an example of a Stairstep Up.

**Life events & stressors:** Between time zero and year 1 her father died, and between year 7 and 8 she separated from her husband and moved. Between year 5 and 6 she had surgery for a health problem that was causing severe pain and during which time she started Paxil. However, she discontinued the antidepressant between years 8 and 9. In summary, the only major changes this woman reported between years 5-6 and 6-7 were surgery because of severe pain, during which time she started antidepressant therapy. It is possible to infer health problems caused the dramatic worsening in her symptoms between those two scores. MTS seemed particularly unrelated to CES-D score or change in score.

Stairstep Down was evidenced by the graph in Figure 38 in a near reverse of the graph in Figure 37.

**Life events & stressors:** This woman moved and married the man she had been living with between time zero and year 1 and improved the nature of her job dramatically, during which time there was no change in CES-D score. The major change this woman

reported between years 1 and 2 was starting psychotherapy. Between years 2 and 3 she lost her job and her father died. Between years 3 and 4 a long-time pet died and her husband had major stresses regarding his business. Health changes occurred between years 2 and 3, when she was hospitalized for worsening of a chronic disease, between year 4 and 5 when she had pneumonia, and between years 6 and 7 when she had a hysterectomy and oophorectomy. She herself reported she was “very depressed” while in the hospital although that occurred between the two low-level annual CES-D scores. Between years 5 and 6 she started St. John’s Wort for depression. This treatment did not prevent the CES-D score rising from 7 to 14 between years 6 and 7.

So far as life events and stressors, this woman had many during her stable low period. Almost all CES-D scores, including the first five, were obtained in January so the Stairstep pattern cannot be attributed to seasonal changes. It is possible to attribute the dramatic improvement in her symptoms between years 1 and 2 to the psychotherapy. MTS seemed particularly unrelated to CES-D score or change in score.

### Combination Patterns

There were a modest number of women who could not be categorized in one of the patterns already listed, who might be characterized as combinations of other subcategories, or perhaps as no identifiable overall pattern. There were no two women with the same combination of patterns. For brevity, only one example will be provided here. The example in Figure 39 could have been labeled Fall & Rise if the final score had not reversed direction and differed so much from the penultimate score.

**Life events & stressors:** This woman included information from her past that might have had an effect on her depressive symptoms as graphed in Figure 39: Her own mother died in midlife when this woman was 15. Fourteen years before entry to the parent study she suffered memory problems after a motor vehicle crash. Four years before entry to the study she was told “in front of my children” that she had a five-year prognosis from a usually fatal diagnosis. She was a mother of seven children and strongly religious. At entry to the study through year 6 until her divorce was final, she reported numerous marital and divorce-related financial crises, including separations and reunions, plus police intervention for domestic violence. Family problems she recorded were physical fight with her son, and his police arrest for his fight with her husband between years 2 and 3, between years 8 and 10 one son reported for truancy, and another son in rehab several times and losing jobs. Her son’s rehab failure was “worse than my divorce.” Her 17-year-old daughter also had a baby she relinquished for adoption between years 8 and 9. Her health events included a cyst on her ovary between years 8 and 9, taking Prozac between years 2 and 5, and St. John’s wort between years 8 and 9. At entry to the study this woman was studying for a teaching certificate after home schooling her own children and working as a teacher’s aide; she then started teaching classes for conduct-disorder children, and finally in year 8 changed to teaching regular classes.

Except for the final CES-D score between years 9 and 10, this woman’s pattern might be explained by the use of an antidepressant during the low-score years. MTS seemed unrelated to CES-D score or change in score.

## SCATTERPLOTS

### Aims

In order to 1) further investigate patterns in the longitudinal data, 2) provide a foundation for transferability of these results in general, and 3) provide a mechanism of connecting the results of the phenomenological with those of the cluster analyses, a final phenomenological analysis was conducted after the determination and labeling of patterns already presented.

### Analytic Procedures

Scatterplots of MEAN00, SLOPE00, AND RANGE00 for women with at least six CES-D scores longitudinally were inspected for potential visual patterns and cut points. Unmarked scatterplots have been presented first for later comparison.

### Results

The scatterplot in Figure 40 of MEAN00 by RANGE00 was similar to scatterplots in Chapter 4 where clusters were identified by different markers (Figures 2, 4, 20, 21, 22). In this case, however, the data were updated for the extra year available from January 2000, and with a slightly larger sample since some women who had had only five CES-D scores through 1999 now had six annual scores available for analysis. Similar to the scatterplots of the cluster analyses, there was a narrow or closed fan shape. The fan shape can be interpreted to mean that women with a low value of MEAN00 generally had a low value of RANGE00. Conversely, the higher the MEAN00 the higher the RANGE00, in general. A second scatterplot of MEAN00 by SLOPE00 (Figure 41) again showed a fan shape. In general, the fan shape indicated that women with a value of

**SLOPE00 close to zero had low MEAN00. A third scatterplot of RANGE00 by SLOPE00 (Figure 42) showed the same fan shape, indicating that at low values of RANGE00 the SLOPE00 was close to zero.**

**Using a visual inspection, groupings within the scatterplots were created and reference lines added. This process was iterative, and the final reference lines were the result after many attempted placements of what was considered the best sense of groups and/or distinctions, both in terms of placement and in terms of numbers of groupings.**

**First, reference lines for MEAN00 were placed at 9, 16, and 22 in the scatterplot of MEAN00 by RANGE00 (Figure 43). These were similar to the phenomenologically determined cut-off values for low, moderate and high in the line graphs. The cut-off at 16 also coincided with the value Radloff first published indicative of major depression (1977). The cut-off at 22 is only two points lower than the high cut-off value of 24 obtained by Roberts et al. (1991) for adolescent girls. While cut-off values for RANGE00 might have been expected to match those for Minor and Major Seesaw, the reference lines that seemed to best distinguish groups of points in the scatterplot were placed at 14 and 27, as indicated in Figure 43.**

**In the scatterplot of MEAN00 by SLOPE00 in Figure 44, reference lines for MEAN00 were retained at the same levels, 9, 16, and 22. SLOPE00 seemed to be best divided into five groups, with cut-off values in the positive slopes at 2.5 and 0.7 and in the negative slopes at -0.9 and -2. The densest grouping of data points were found in the group with the lowest MEAN00 and the SLOPE00 closest to zero. The next most dense grouping of data points was seen in the same area of SLOPE00 near zero, but at the next**

higher grouping of MEAN00 from 9 to 16. These groupings repeated the results of the phenomenological analysis of line graphs, where Stable and Changing No Trend categories contained the most common patterns.

In the scatterplot of RANGE00 by SLOPE00 (Figure 45), cut-off values for SLOPE00 were not exactly the same as in the scatterplot of MEAN00 by SLOPE00 in Figure 44. While the number of cut-offs were the same, the cut-off for the negative slopes were placed at  $-1.2$  and  $-0.7$  and for the positive slopes  $0.2$  and  $1.2$ . The data points were again quite dense around the slope of zero, with a more skewed distribution toward the negative slopes. A highly unequal division of the negative slopes between the rare, sparse values greater than  $-1.2$  and those between  $-0.7$  and  $-1.2$  was therefore created. The relatively large group of data points at lowest and middle RANGE00 between the cut-off values of SLOPE00 at  $.2$  and  $1.2$  was reflective of the moderately large group of women with Increasing Oscillation Rising, as well as Minor and Major Seesaw rising.

### Summary

Patterns were divided into waveform, non-waveform, and combinations. Waveforms had three important characteristics: amplitude, frequency, and level. Level had two dimensions, level and trend. Waveform patterns of this data set demonstrated most of the possible combinations of amplitude, frequency, and level. Four levels were distinguished: non-depressed, low, moderate, and high. The single largest group of women were stable at their own level, overwhelmingly the non-depressive level, but also a few at low and moderate levels.

There were also women who fluctuated around their own level. Additionally, some women fluctuated around a rising level and some around a falling level. A fairly large group of women who became more stable over time while they generally improved (Decreasing Oscillation Falling). Another fairly large group of women became less stable while they generally worsened (Increasing Oscillation Rising). Another fairly large group of women also became less stable each at her own level without an overall trend to change (Increasing Oscillation No Trend); nearly as large a group became more stable each at her own level without an overall tend to change (Decreasing Oscillation No Trend). Again, in mirror fashion to the rising category, other subcategories of falling patterns (negative SLOPE00) were not represented by large numbers of women. Patterns had their own period or frequency of fluctuation over the years, and rarely they became shorter, or longer.

The non-waveform patterns were labeled Stable or flat, Generally Linear Falling, Generally Linear Rising, Blip Up, Blip Down, and occasional bad or good year. Combination patterns could not be classified in any of the above subcategories. Several subcategories, such as combination patterns, Blips, and especially Increasing and Decreasing Oscillation tended to demonstrate change in the pattern itself over time. There were only a modest number of women in the Blip and Stairstep categories (fewer than 10 combined). Life events and stressors and MTS, including use of hormone replacement therapy, were not consistently related to CES-D score or changes in score. Antidepressant use was also not consistently related to CES-D score or changes in score.

## Discussion

### Interpretation Of Findings

In light of previous studies of longitudinal depressive symptoms in midlife women (see Chapter 2), the finding that a plurality of women never demonstrated symptoms of any clinical significance was not surprising. Many proposed concepts about the heterogeneity of depression (subsyndromal, dysthymia, bipolar, recurrent brief, minor, chronic, double) were also consistent with current findings.

For example, subsyndromal depression might be expressed by CES-D scores at levels less than 16, such as were described as moderate or low in the current study. Dysthymia might also be evident in stable scores with no trend at low and/or moderate levels. A Major Seesaw pattern might be expected to reflect what one would expect for women who have untreated bipolar disorder. However, epidemiological studies have consistently found bipolar disorder to be prevalent in only about 1-2% of a population (Kessler et al., 1994; Weissman et al., 1996). From the sample of 205 women, there should be approximately two to four women. Yet there were at least twice as many women with a Major Seesaw pattern. Also, the fewer than five women in the sample known to have bipolar disorder were all being treated. It was unlikely, therefore, that bipolar disorder was a significant predictor of the Major Seesaw pattern, whether Stable, Rising or Falling. The annual point assessments of yearly CES-D scores might be capturing by chance fluctuations in depressive symptoms that are actually more frequent. For example, RBD (Angst & Hochstrasser, 1994) might appear as Stable No Trend, Seesaw, or even Blip patterns.

**Patterns labeled Seesaw, Slow Oscillation, Rise & Fall, Grand W, Blips**

Up/Down, or occasional good/bad year might be expected to support the causal models of depressive episodes as reactions to adversity. While no simple consistent relationships between life events and depressive symptoms were noted in this study, no competing explanations for the fluctuating patterns were identified either.

Women whose scores showed patterns with rising slopes are of greatest clinical concern. The single largest subcategory within Rising was Increasing Oscillation Rising. This pattern implied that increasing CES-D scores (worsening depressive symptoms) did not usually proceed simply but were part of an increasing general variability or increasing overall instability. Conversely, the even slightly larger group of Decreasing Oscillation Falling scores (improving symptoms) meant that women generally became more stable as symptoms improved.

There might be other combinations of fluctuation, level (both mean and trend), and duration (longitudinal course) as well as changes over time in all three of these dimensions and their interactions that have clinical importance, although the particular clinical implications of many of these patterns is not known at this time. However, the observation that many women had at least some symptoms more often, and that symptoms fluctuated more often, than many primary care providers might be aware of (Judd et al., 1997) could lead to improved monitoring and interventions for women.

While the DSM-IV did add the option of course modifiers to diagnoses of mood disorders, the issue of longitudinal course of depressive symptoms might be more significant and more complex than is currently known. Further diagnostic modifiers

might also be appropriate, especially if severity, duration, and long-term pattern were considered separately. Lability of depressive symptoms as seen in Minor and Major Seesaw patterns could possibly require different interventions than stable high patterns, particularly in terms of antidepressant drugs or psychotherapy. Researchers might also be able to investigate predictors and even triggers for changes given a better understanding of the complexity of longitudinal patterns.

Some new findings included the concept of fluctuation around a stable level, especially repeated as a pattern at low, moderate, and high levels. For the same patterns to be replicated at several levels was unexpected. If women from patterns labeled Seesaws, Rising, Falling, and Rise & Fall are combined, then it could be said that a majority of women fluctuated in severity of symptoms over the years. Fluctuations occurred not only in severity of symptoms but also in period (Increasing or Decreasing Oscillation) and trend. The concept of increasing stability (Decreasing Oscillation), whether at a stable level, rising level, or most commonly falling level was also new. The mirror pattern of decreasing stability (Increasing Oscillation), most commonly at a rising level, also has not been identified previously, to the author's knowledge.

#### Comparison With Other Studies

One other study that conducted a longitudinal investigation of more than two time points measuring CES-D scores in midlife women was the Manitoba Project on Women and their Health in the Middle Years (Kaufert et al., 1992). These researchers used the CES-D score as a dichotomous variable and repeated its measurement at five time points each six months apart. They presented longitudinal patterns of the dichotomous variable

in boxes, black for total CES-D score  $\geq 16$  and white for  $< 16$ . They found that “one sequence (namely, a score of less than 16 at every interview) is common to more than 74%” or 347 women (p. 148). Their results were compatible with results of the current analysis if the current subcategories at low levels ( $< 16$ ) were combined with the stable non-depressed pattern.

Kaufert and colleagues also reported that cross-sectionally

the percentage of women with a score of 16 or above on the CES-D scale was relatively stable across all five interviews [every six months for 2½ years], ranging between 9% and 11%, but membership in the group of depressed women varied from one interview to the next (p. 148).

While the latter could be compatible with Seesaw patterns, the percentages of women with CES-D scores  $\geq 16$  every year in the SMWHS study were 22-24%. The discrepancy between 22-24% in Seattle and 9-11% in Manitoba might have been due to the demographic differences between samples in employment, education, and income (see Chapter 2). Further, the Manitoba researchers wrote that of the women who did have at least one score  $\geq 16$ , “most patterns are particular to no more than 3 or 4 women” (Kaufert et al., 1992, p. 148). A pattern of only one of the scores being high occurred for 60 Manitoba women, 13% of their total sample and was therefore their second most populous type of pattern, populous patterns having more than 10 women each.

While their narrative interpretation of their results gave the impression of transient depression for that minority that were ever depressed, in light of the current study their results bear a more complex interpretation. Sixty-two or 13% of Manitoba women also had intermittent, recurrent, or persistent high CES-D scores, that is, three or

more high scores of the five total. Their results could be consistent with those of the current study if high level Major and Minor Seesaw and Increasing and Decreasing Oscillation patterns were combined. In spite of their descriptive results, however, Kaufert et al. (1992) reported that they had difficulty interpreting the 5-point patterns (see Chapter 2).

Chen et al. (2000) also studied patterns of depressive symptoms considering level and course at two time points 13 years apart, with lifetime number of depressive episodes and the intervening 13-year course determined by recall. Stressful life events included children moving out but not menopause or the menopausal transition. They identified four levels, none, mild, moderate, and severe depression. The latter two both qualify as MDD, however moderate evidences five to six of the DSM-IV criteria, and severe seven to nine. Although Chen and colleagues did not distinguish gender, they found a mean age of lifetime first episode in the early 30s for each level, which implies that as many people experienced their first depressive episode in their 40s as in their 20s. Latent variable analysis yielded five groups including non-depressed. Course of depressive symptoms was similar in four of the five groups. However, in the group with lowest level of symptoms the mean age of first episode was 43 years and the lifetime number of episodes was only one, both significantly different from the other four groups of depression severity. Gender displayed a significant interaction with risk factors in this lowest severity group, where stressful life events were significantly more likely in women than in men. In the second highest severity group women were less likely than men to have a family history of depression. These results are consistent with the Rise &

Fall and perhaps Blip Up patterns identified in the current study. These researchers also found a pattern of severity that changed from mild to moderate to mild and another pattern of moderate to severe to moderate, both of which could be consistent with Seesaws at two different levels.

#### Comparison Of Phenomenological Results To Cluster Results

In an earlier chapter, cluster analysis was used to identify longitudinal patterns in CES-D scores (see Chapter 4). Comparison of the results of the phenomenological analysis to the results of the cluster analyses in Chapter 4 allowed further understanding of longitudinal patterns of CES-D scores in midlife women. Many patterns were similar across results of both analyses.

The single most common pattern remained the stable non-depressive level in both phenomenological and cluster analyses. This meant that the plurality of women, stable non-depressed, were consistently identified regardless of the analytic approach. In the cluster analyses this was K-Cluster #7 and in hierarchical clustering it was #1, the one with non-depressive MEAN99, low RANGE99 and SLOPE99 of zero.

The stable Minor and Major Seesaw patterns of the phenomenological analysis repeated the K-Clusters # 2 and # 1, respectively, with low or moderate MEAN99, higher RANGE99 and still SLOPE99 of zero. While not always discriminated in terms of cluster membership in the hierarchical cluster analyses, in the k-clustering there was also a group of women with fairly stable trends over time at a low to moderate level yet who had large variability, in essence a Seesaw pattern. Rising and Falling categories were again patterns present in the cluster analyses as K-Cluster # 8 and #3, and hierarchical

Clusters #3 and #2, respectively, where SLOPE99 was positive and RANGE99 was moderate to large, or SLOPE99 was negative and RANGE99 was mod to large.

The pattern described as Rise & Fall was also identified in some of the cluster analyses. Slow Rise & Fall could be evident in K-Cluster #6, although the number of women identified phenomenologically with this pattern was smaller than the population of the k-cluster. A small number of women with high levels, high variability and generally worsening trends were consistently identified as Major Seesaw Rising High or K-Cluster # 9 and hierarchical Cluster #4.

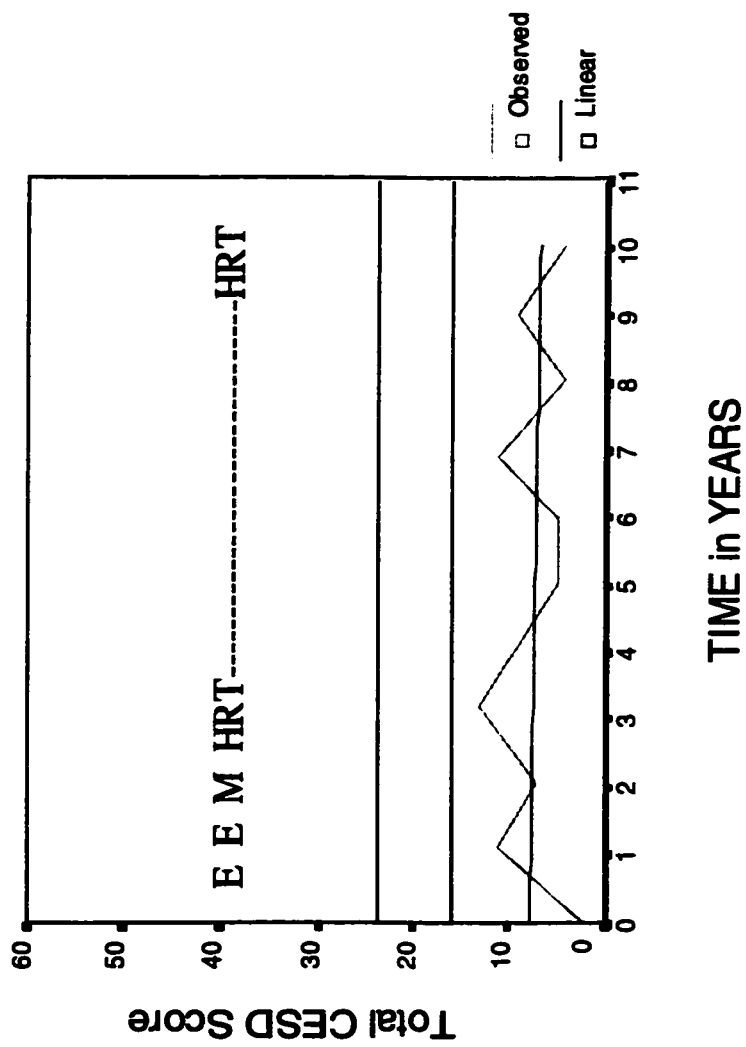
The rare phenomenological patterns, such as Generally Linear Falling, Generally Linear Rising, Grand W, Blips (early, middle, or late) and occasional bad or good year, were not well identified by the cluster analyses. They were often cluster analysis outliers. In the phenomenological analysis there were different outliers than in the cluster analysis, the combination patterns. In the cluster analyses, combination patterns were more likely to have their strongest characteristic, such as mean or range, allow for clustering with other women's data. Blips, Slow Oscillation, and especially Increasing and Decreasing Oscillation at whatever SLOPE00, were only identified through phenomenological analysis. Common statistical techniques and even hierarchical linear modeling would not have been able to create these various longitudinal patterns.

#### Limitations

While the CES-D score was determined at a single point in time, the line graphs created a connecting line between each consecutive point. Even where there were missing years and the line was interrupted, it was noted that the eye interpolated a linear

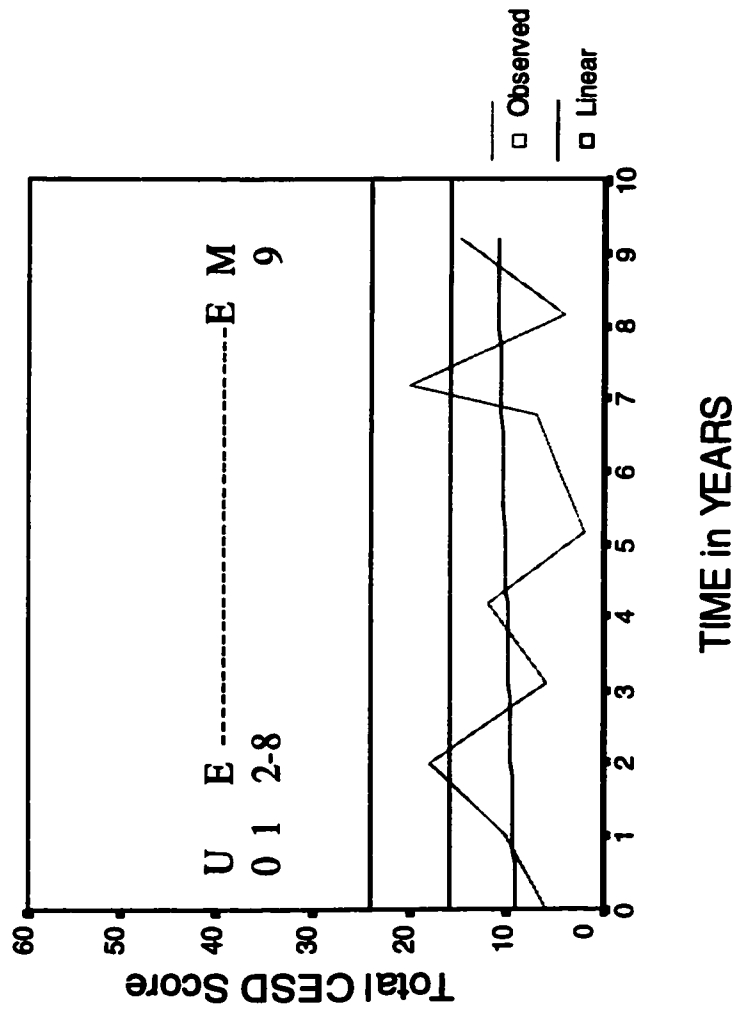
connection between any two consecutive points. While this facilitated pattern labeling it also created a sense that depressive symptoms were scored continuously, which they were not. It was possible that symptoms fluctuated considerably between any two measured time points, or that the change from one score to the subsequent one was not a smooth linear one but perhaps seesaw or stairstep.

Life events and stressors might not have been identified in an optimal fashion. Although it was expected that for a phenomenological analysis a prospective free form of recording would be ideal, the lack of any consistent relationship with CES-D scores was unexpected. It is possible that if standardized measures of life events and stressors had been used, for example the Life Events Survey adapted by Norbeck (1984) and measured for occasional years of the SMWHS, a consistent relationship between stressful life events and depressive symptoms would have been found. Daily ratings of depressed mood and other depressive symptoms were also available intermittently from diary recordings and could be included in future analyses.



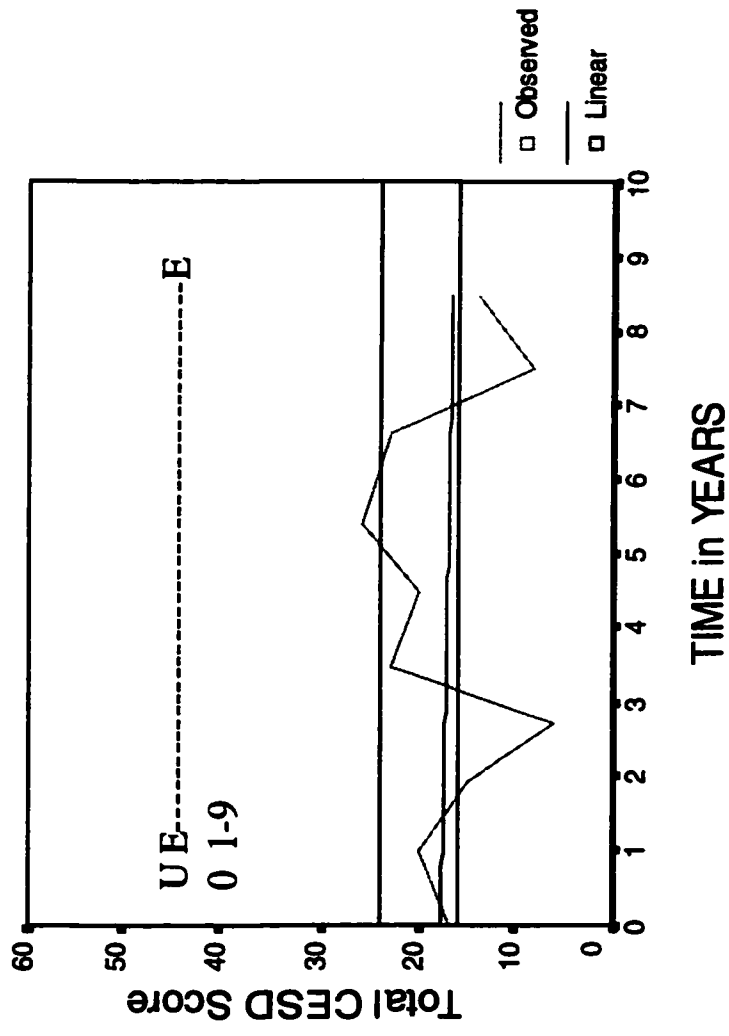
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT use; OC=OC use; O=Other; U=Unknown

Figure 23  
Minor Seesaw Low No Trend



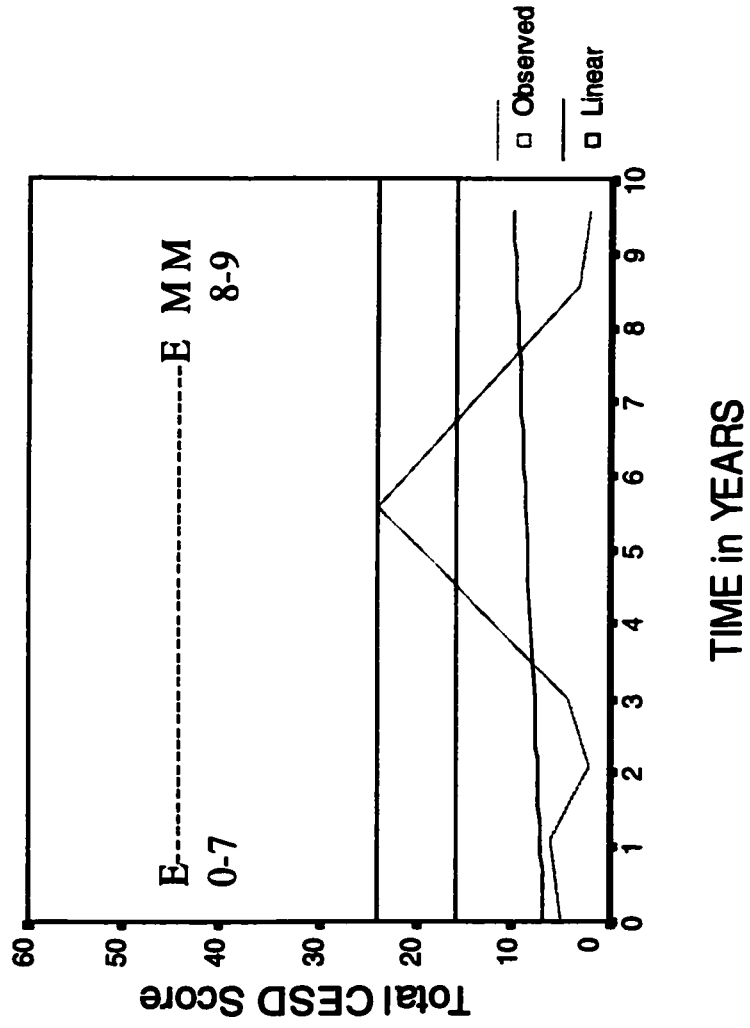
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT use; OC=OC use; O=Other; U=Unknown

Figure 24  
Major Seesaw Low No Trend



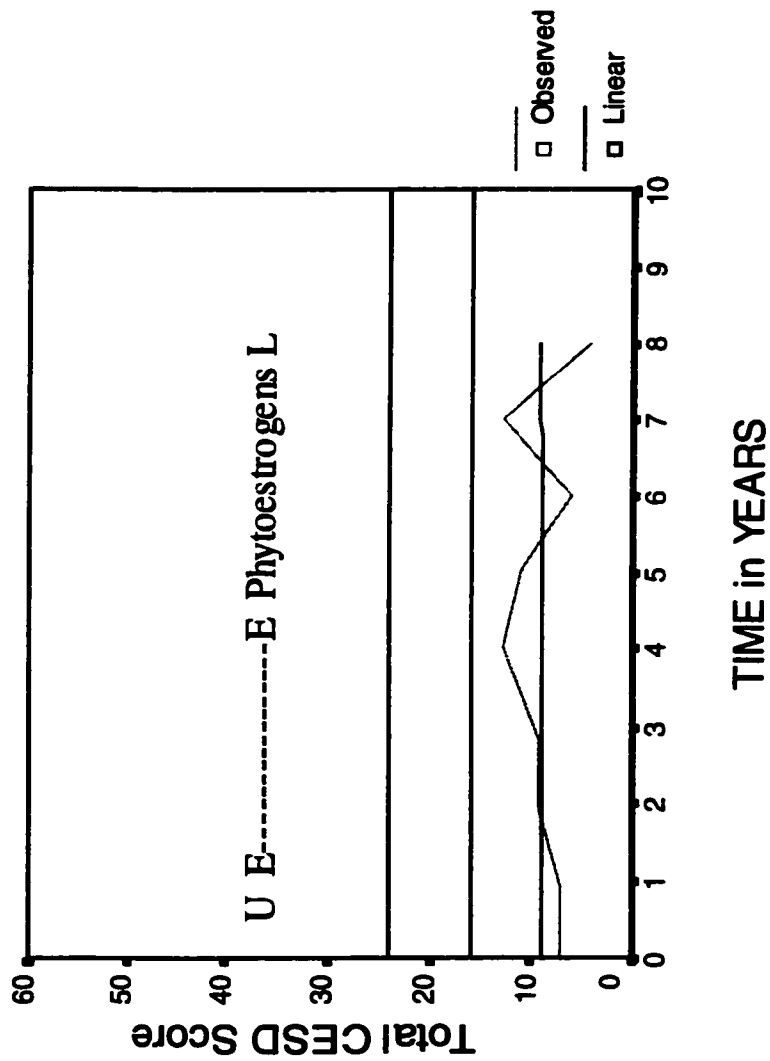
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT use; OC=OC use; O=Other; U=Unknown

Figure 25  
Slow Oscillation High No Trend



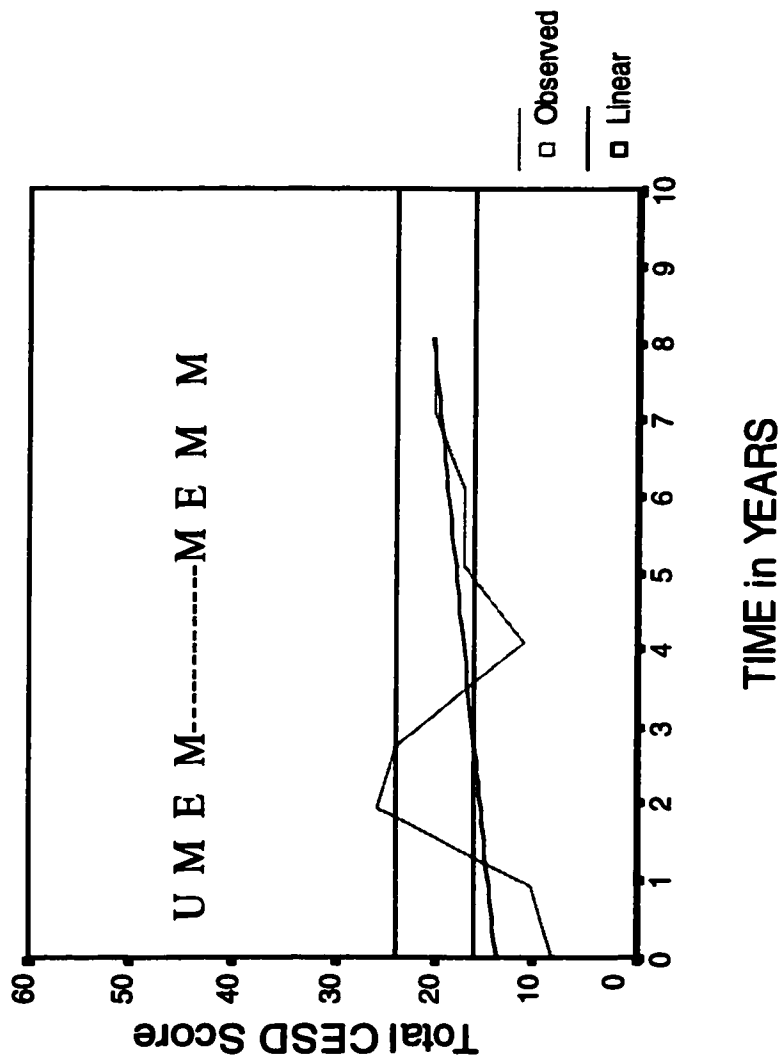
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT use; OC=OC use; O=Other; U=Unknown

Figure 26  
Rise&Fall



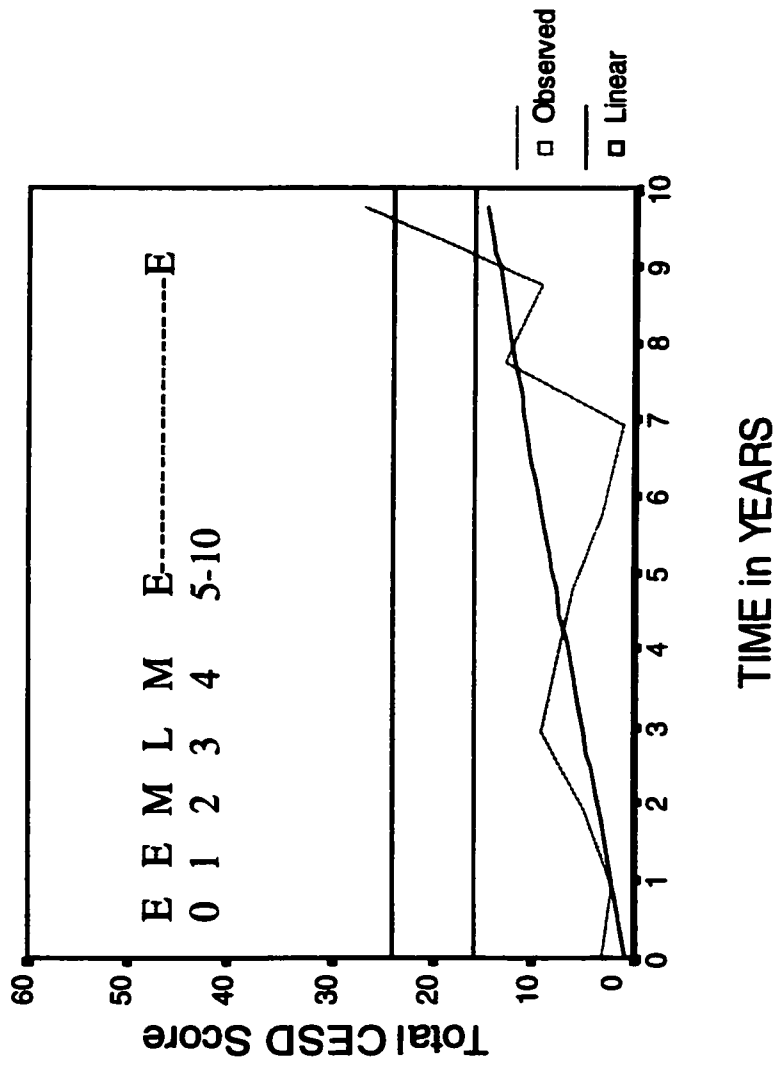
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT use; OC=OC use; O=Other; U=Unknown

Figure 27  
 Increasing Oscillation Low No Trend



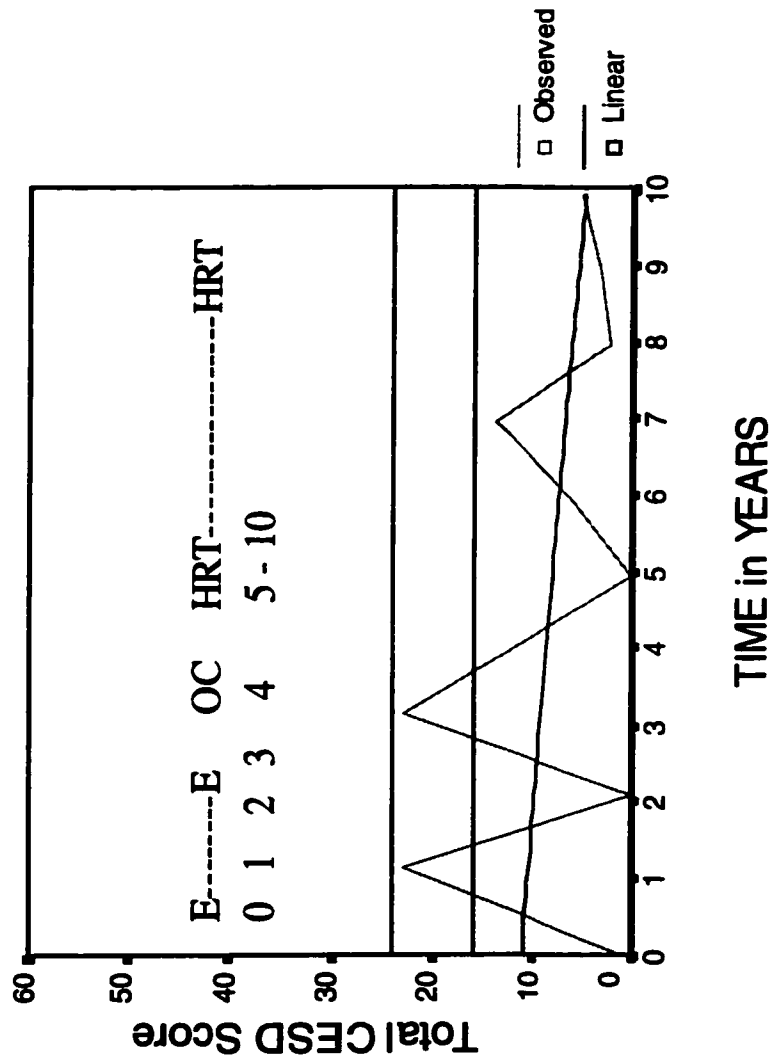
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT; OC= OC use; O=Other; U=Unknown

Figure 28  
 Decreasing Oscillation High No Trend



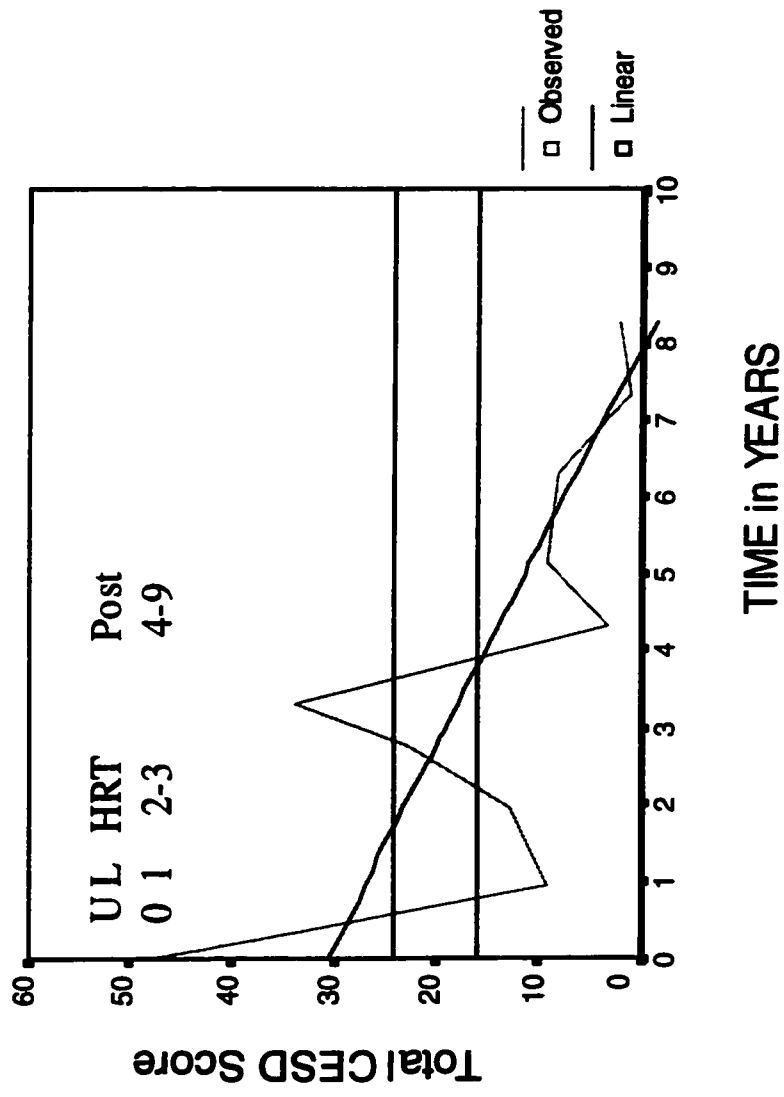
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT; OC= OC use; O=Other; U=Unknown

Figure 29  
Increasing Oscillation Rising



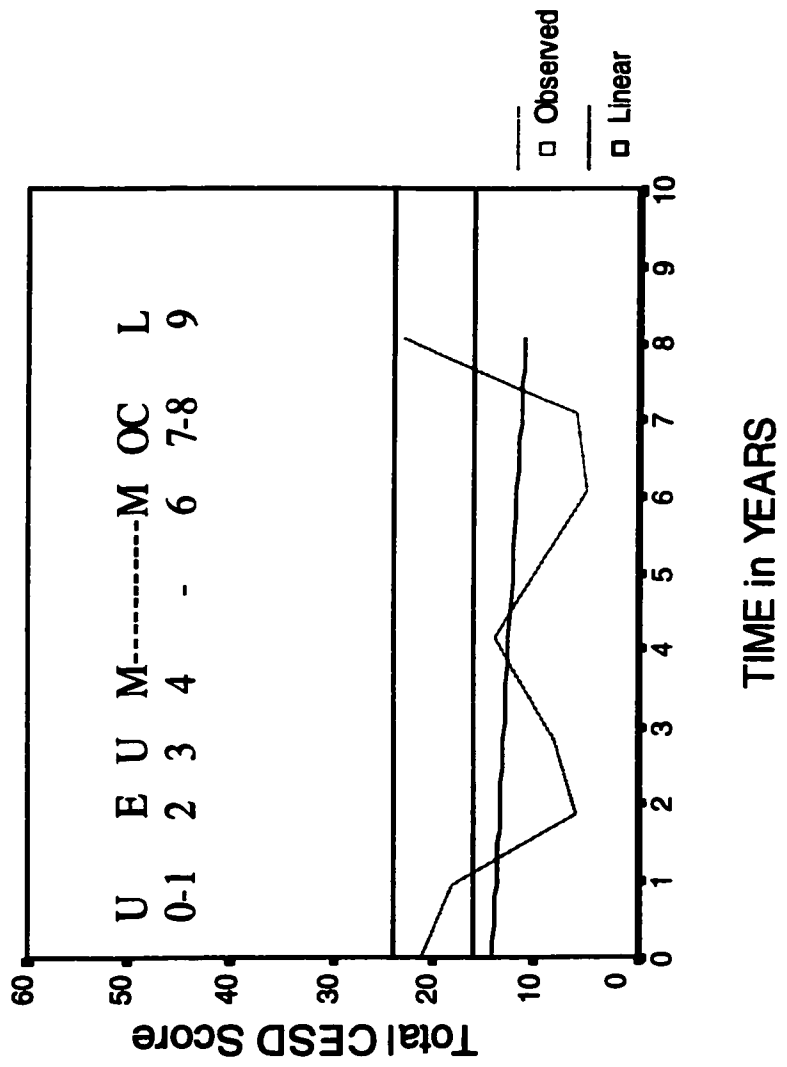
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT; OC= OC use; O=Other; U=Unknown

Figure 30  
Decreasing Oscillation Falling



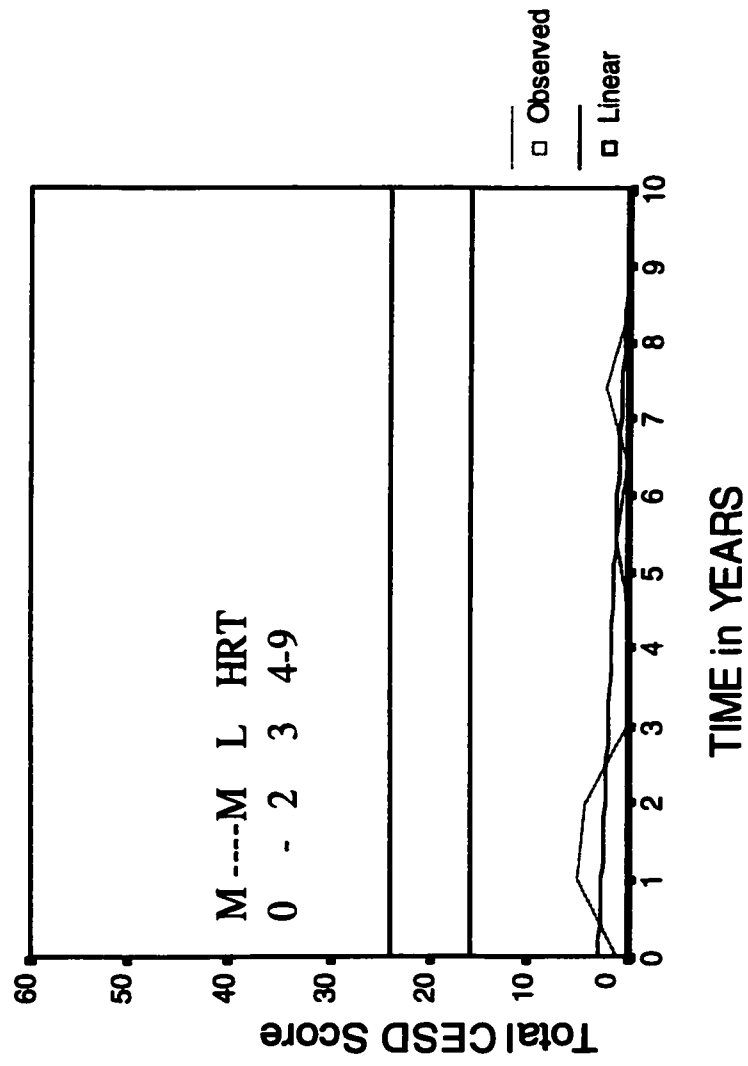
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT; OC= OC use; O=Other; U=Unknown

Figure 31  
Decreasing Oscillation Maximally Falling



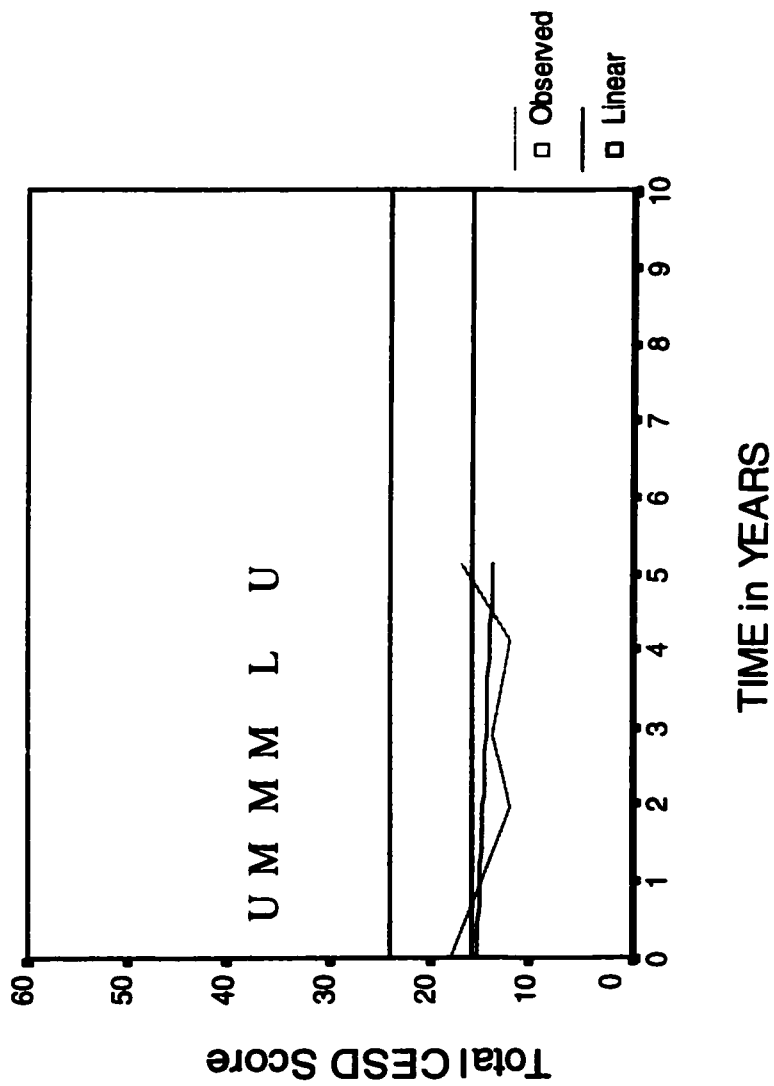
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT; OC=OC use; O=Other; U=Unknown

Figure 32  
Grand W



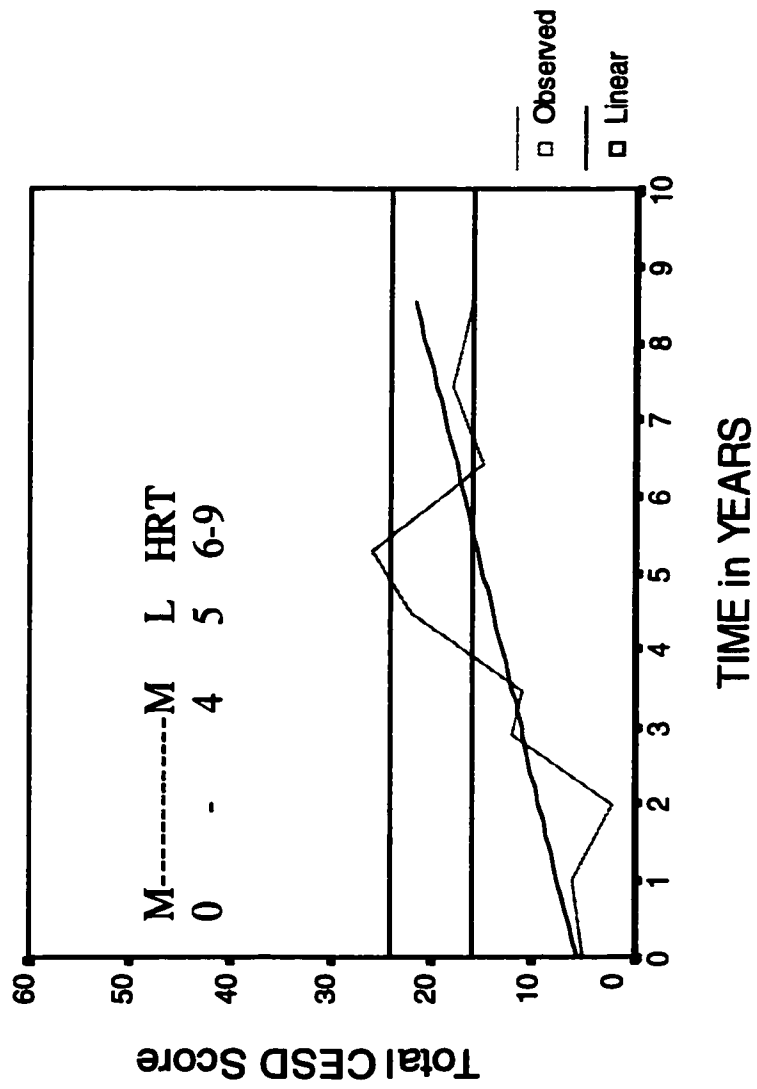
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT; OC= OC use; O=Other; U=Unknown

Figure 33  
Stable Non-depressed



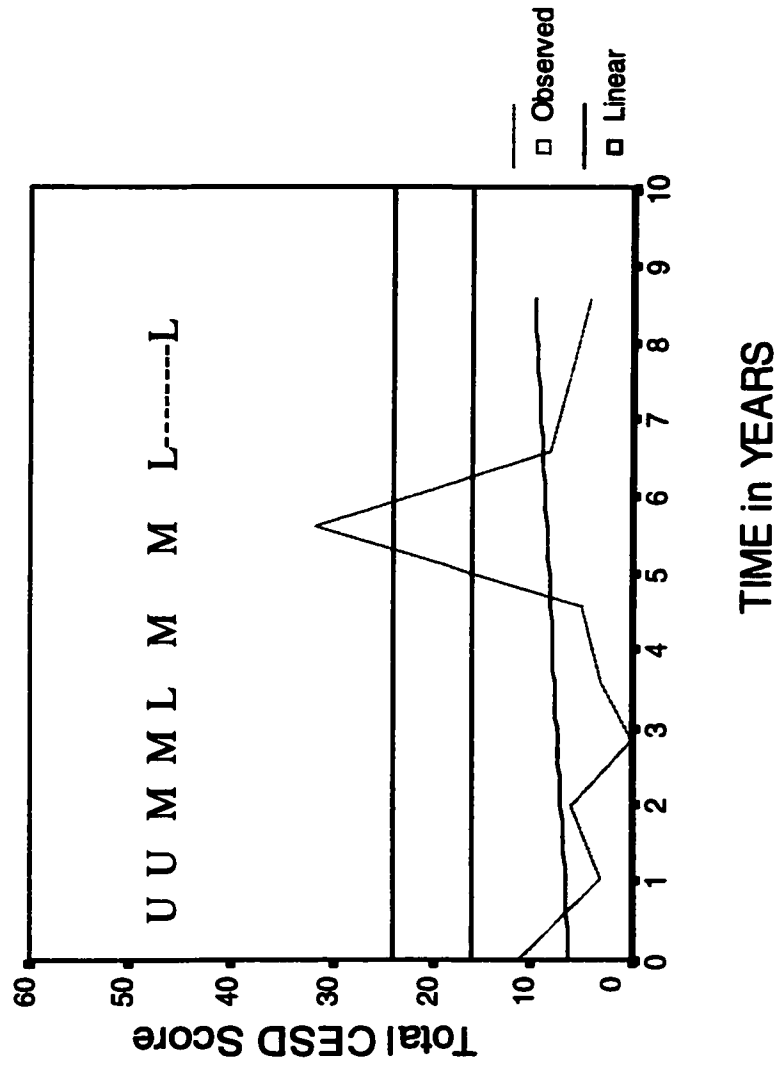
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT; OC= OC use; O=Other; U=Unknown

Figure 34  
Stable Moderate



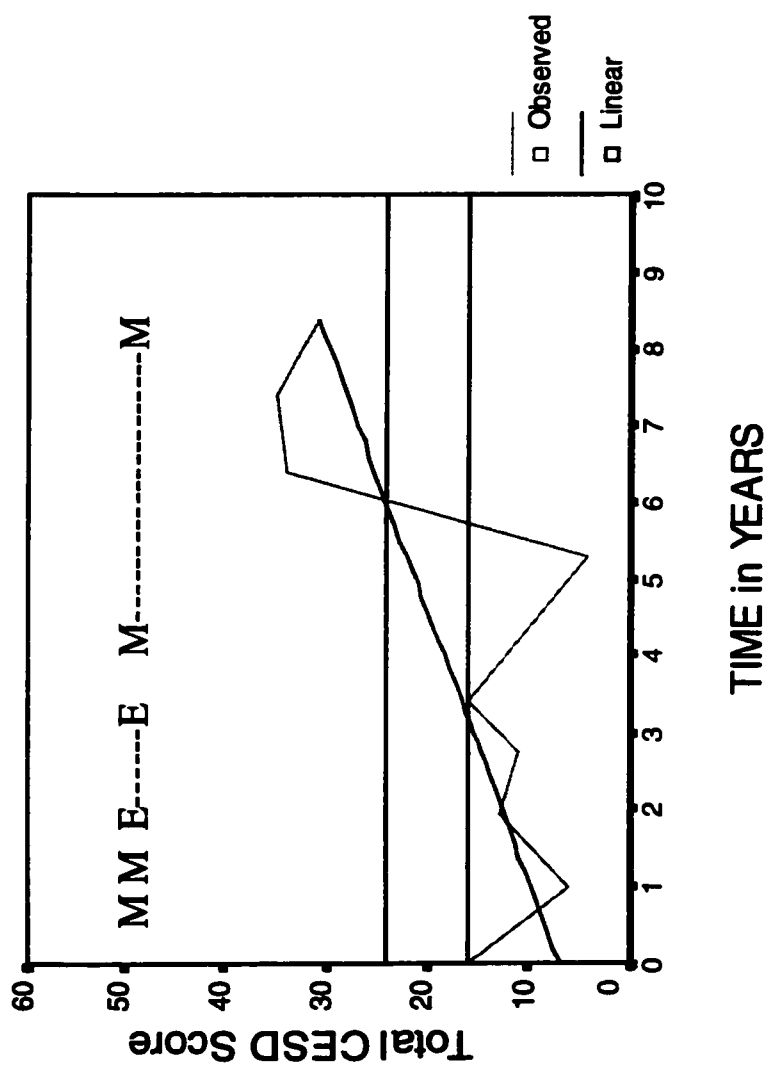
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT; OC= OC use; O=Other; U=Unknown

Figure 35  
Generally Linearly Rising



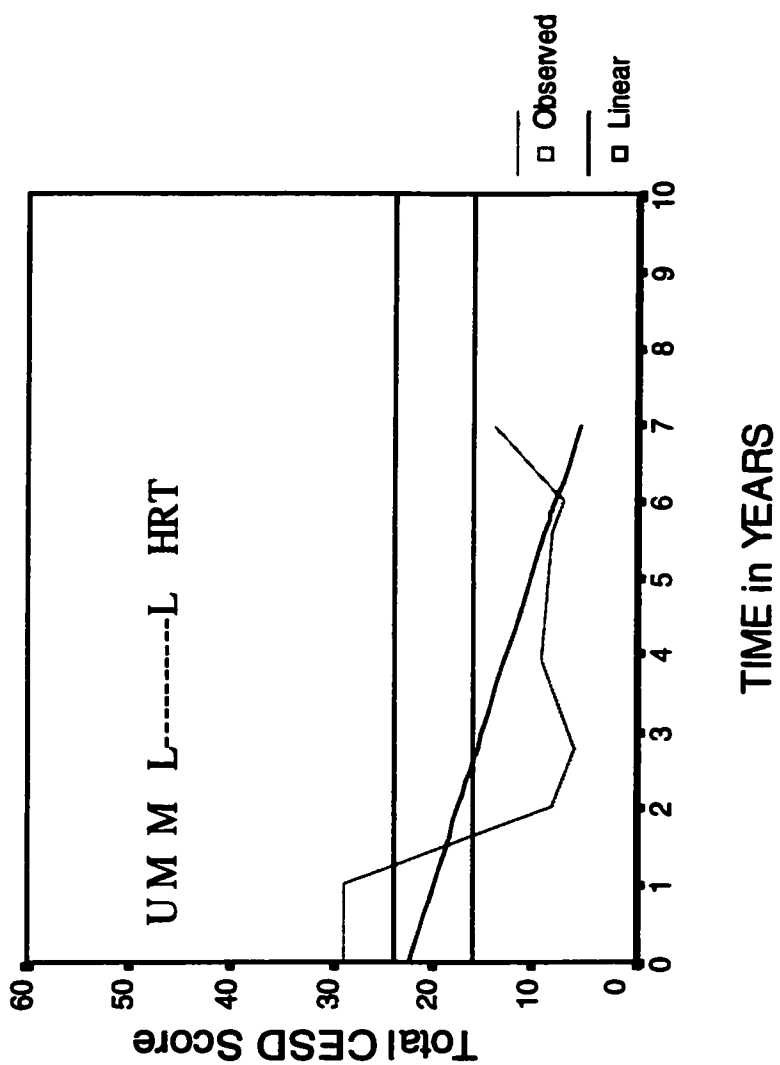
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT; OC= OC use; O=Other; U=Unknown

Figure 36  
Blip Up Late



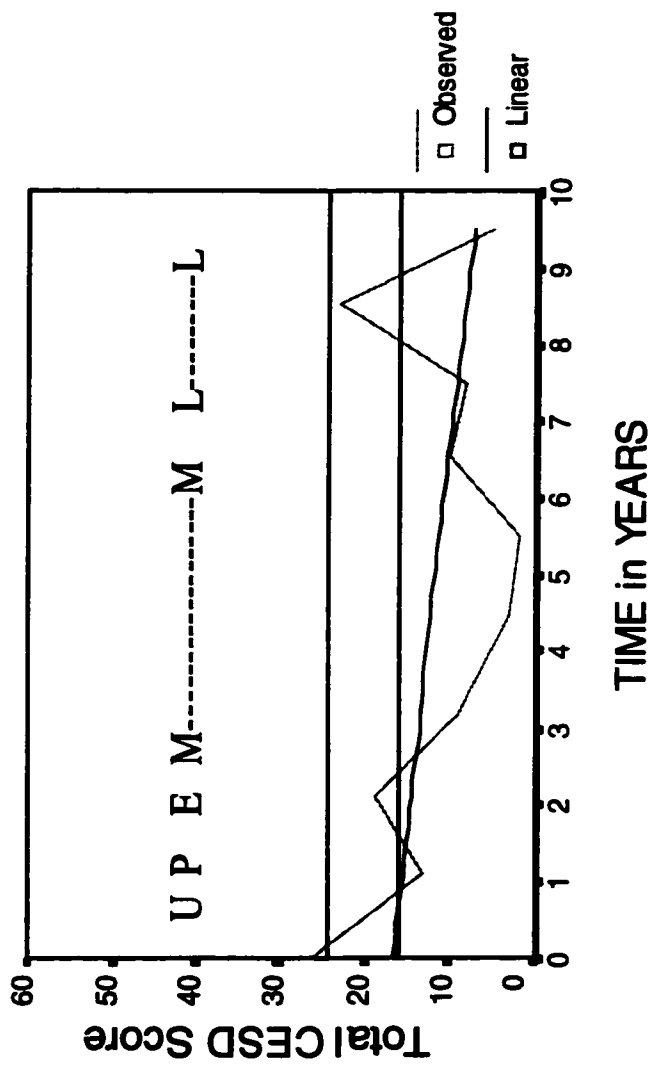
P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT; OC=OC use; O=Other; U=Unknown

Figure 37  
Stairstep Up



P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT; OC= OC use; O=Other; U=Unknown

Figure 38  
Stairstep Down



P=Pretransition; E=Early; M=Middle; L=Late; PM=Postmenopausal; H=HRT; OC= OC use; O=Other; U=Unknown

Figure 39  
Combination: Decreasing Oscillation Falling & Increasing Oscillation

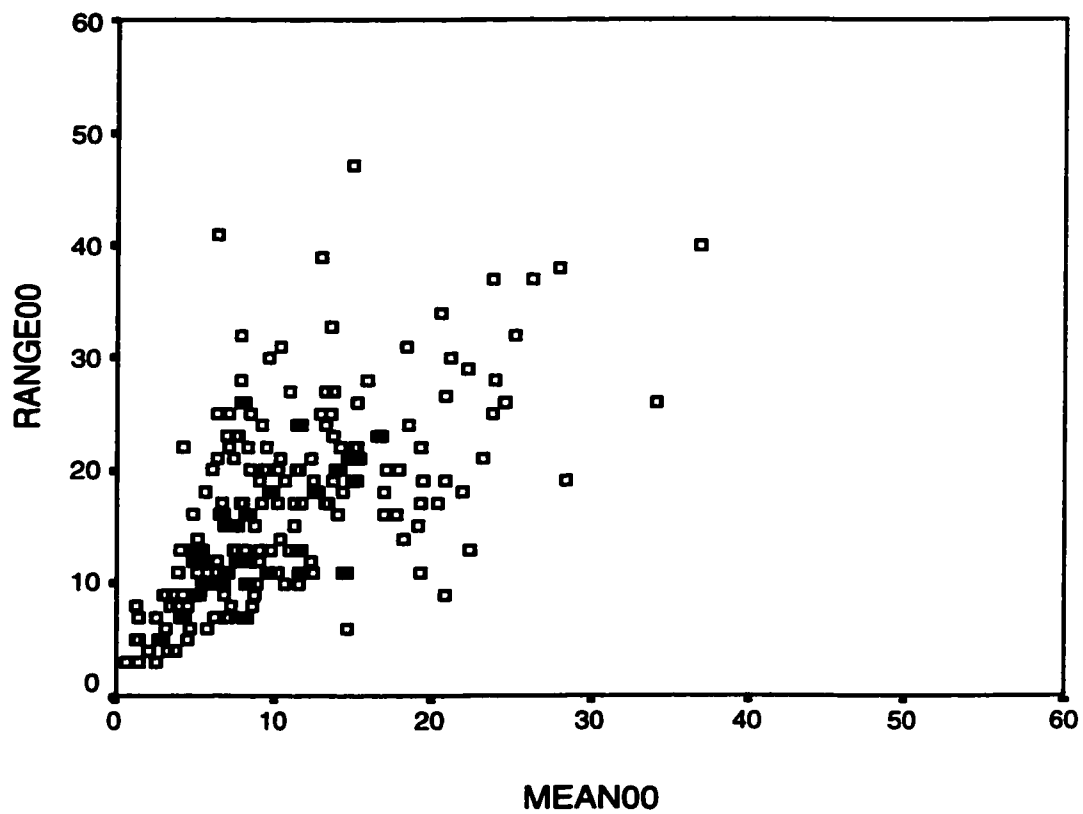


Figure 40  
Scatterplot of MEAN00 by RANGE00

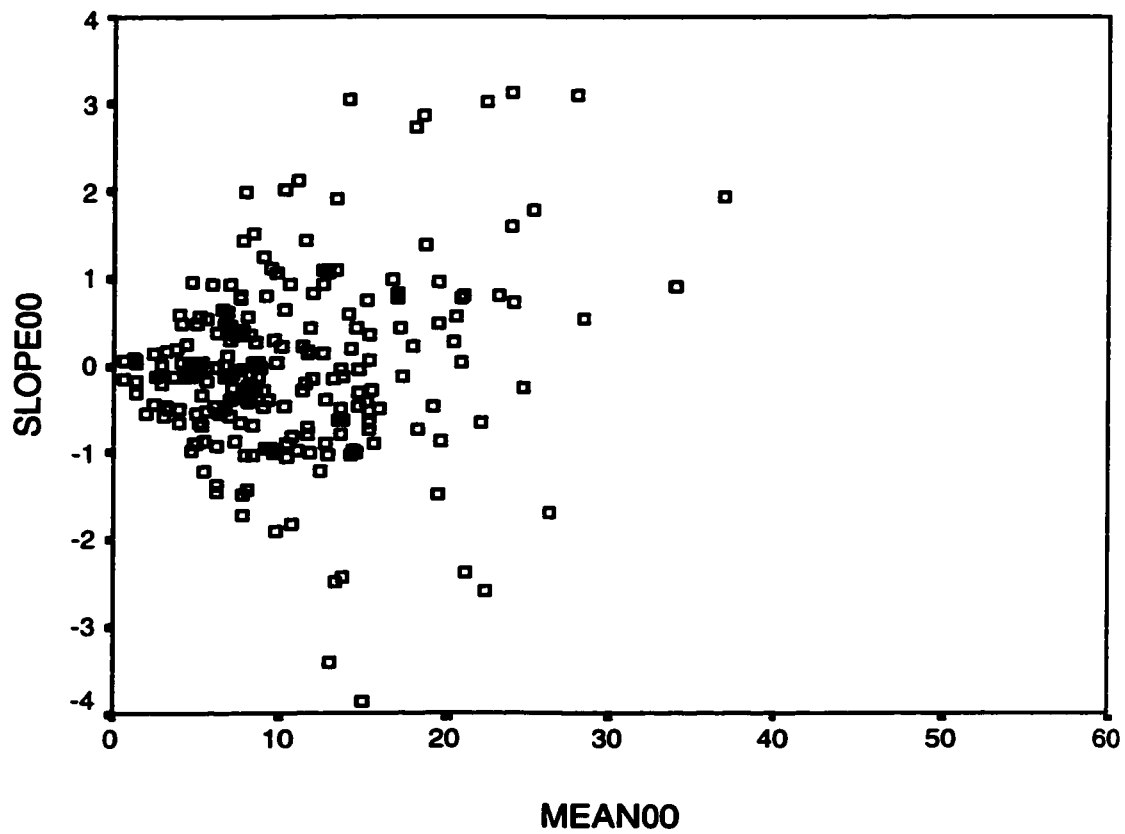


Figure 41  
Scatterplot of MEAN00 by SLOPE00

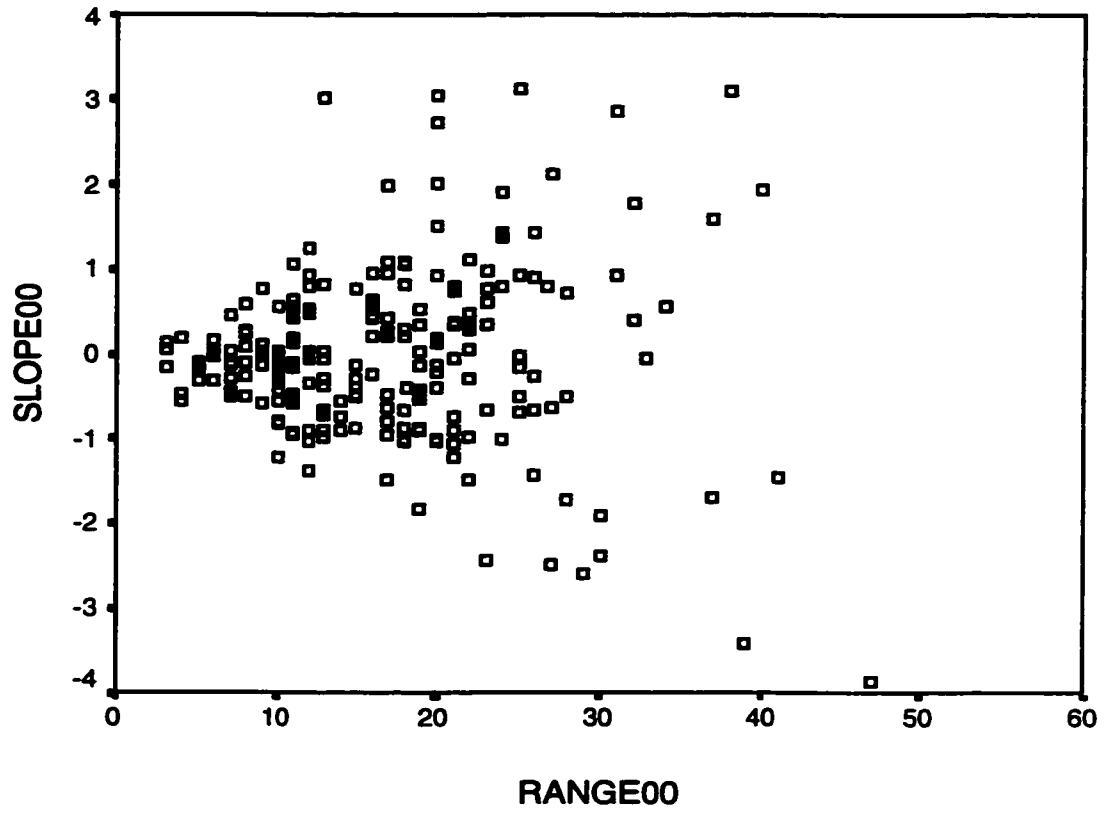
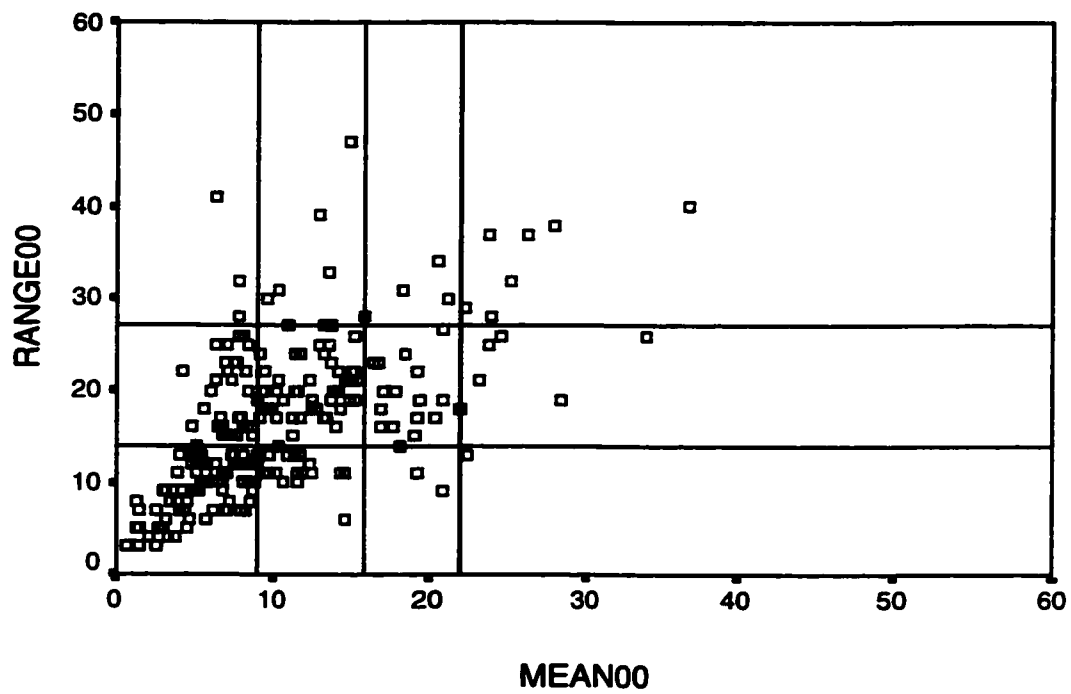


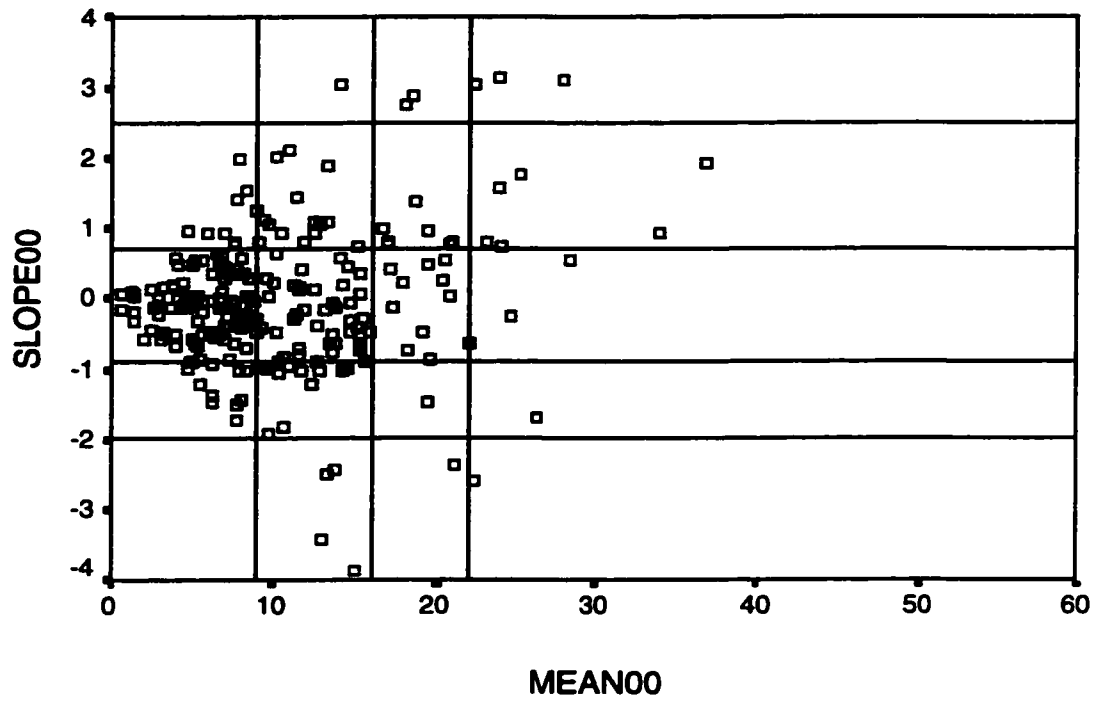
Figure 42  
Scatterplot of RANGE00 by SLOPE00



MEAN00 reference lines at 9, 16, 22

RANGE00 reference lines at 14, 27

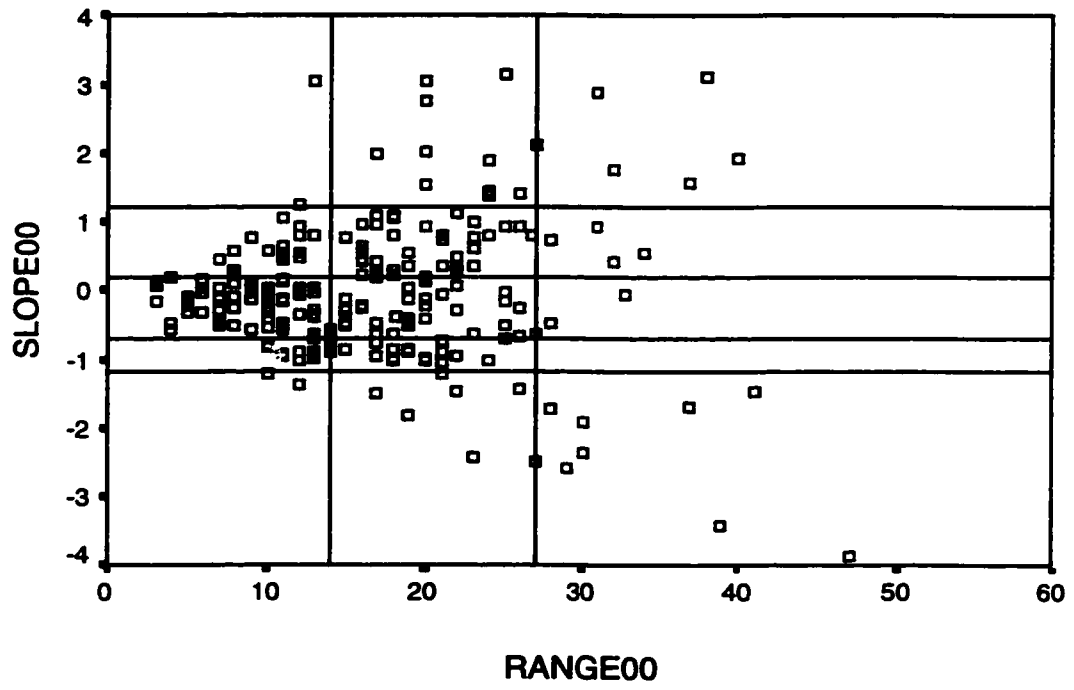
Figure 43  
Scatterplot of MEAN00 by RANGE00  
Reference Lines Added



MEAN00 references lines at 9, 16, 22

SLOPE00 reference lines at 2.5, .7, -0.9 -2

Figure 44  
Scatterplot of MEAN00 by SLOPE00  
Reference Lines Added



RANGE00 reference lines at 14, 27

SLOPE00 reference lines at 1.2, .2, -.7, -1.2

Figure 45  
Scatterplot of RANGE00 by SLOPE00  
Reference Lines Added

## CHAPTER 6: GENERAL DISCUSSION

The purpose of this chapter was to discuss results in general from both analyses, including general comparisons to published studies, limitations, clinical implications for nurses and nurse-practitioners, and implications for further research.

### Interpretation Of Findings

Certain models of depression were supported by the results of the both the cluster and phenomenological analyses. The concept of a stable trait as an important determinant of longitudinal patterns of depressive symptoms (or their absence) (Duncan-Jones et al., 1990; Ormel & Wohlfarth, 1991) was consistent with the general stability of CES-D scores of most clusters and patterns in the current study. For the most part, each woman had her own level at which she was either stable or around which she fluctuated. This stability was demonstrated by the majority of slopes being near zero, that is, the majority of patterns showing no trend. Stability was especially prominent in patterns that were stable with no fluctuations (low range), most of these patterns at the non-depressed level.

A particular contribution of the current study was identification of repetition at different levels of the same patterns, such as Minor Seesaw No Trend at both low and moderate levels, or Cluster #1 and Cluster #6 showing similar patterns at low and moderate levels, respectively. Another contribution of this study was identification that most women with a level above non-depressed demonstrated variability over the years; that is, women's depressive symptoms were more likely to change from year to year than to remain constant. However, many women had stable patterns of change. For some it

was modest change and for others large changes from year to year to year. The lability of depression for some women already has been intimated in the diagnoses of both RBD and PMDD.

Patterns from both cluster and phenomenological analyses labeled Rise & Fall/Fall & Rise might be expected to support the causal models of depressive episodes as reactions to adversity. While no consistent relationships between life events and depressive symptoms were noted in this study, no competing explanations for the fluctuating patterns were identified either. It remains to be seen whether a complex model of stable personality characteristics interacting with change-inducing life events is predictive of depressive symptoms at a particular time point.

Similar to most of the longitudinal research on depression in midlife women (see Chapter 2), the current study found no relationship between depressive symptoms and MTS or menopause in any of the analytic approaches. Also similar to other longitudinal research on depression in general, this study found that the best predictor of midlife depression in women was a history of previous depression, that is, the majority of depressive symptoms were recurrences and/or exacerbations. In particular, history of sexual assault/abuse was related to chronic high levels of depressive symptoms. Unlike some studies (Hällström & Samuelsson, 1985; Kaufert et al., 1992; Woods & Mitchell, 1996) this study found inconsistent relationships of depressive symptoms to recent or current life events.

In terms of level or severity, Chen et al. (2000) have proposed that depression is a more continuous health problem with a less well-defined clinical threshold than

previously understood. In freeing the current study's definition of depressive symptoms from threshold-defined categories it was possible to more thoroughly investigate Chen et al.'s proposal. Chen et al.'s (2000) description of four levels, non-depressed, mild, moderate, and severe, might correspond to the levels found in this study of non-depressed (CES-D score around 4-7), low (score around 9-11), moderate (around 15-17), and high (over 20). Lewinsohn & Teri (1982) also found that "trichotomizing CES-D scores" into less than or equal to 8, 9 to less than 17, and 17 or greater provided fewer false negatives or false positives compared to a diagnosis based on clinical interview. It might be tempting to label the current study's moderate or low levels as minor depression. However, while the presence in this data set of a distinct group of women with low and another with moderate levels of CES-D scores, whether stable or fluctuating, was intriguing, the relationship between diagnoses or clinical significance and low or moderate level CES-D scores has not been established.

In terms of longitudinal pattern, Judd et al. (1997) found that among people identified from a general population with depressive symptoms, one year later the majority, 71% still had symptoms, consistent with the majority of women who had longitudinal patterns with no trend. They also found "enormous amount of flux and change observed in these respondents during only 1 year of follow-up" (p. 14), consistent with seesaw patterns and other patterns of lability.

What these findings mean for the researcher and clinician is continued evidence that depression is not a homogeneous monolithic health problem but a complex spectrum of symptoms (Chen et al., 2000; Judd, 1997a, 1997b; Winokur, 1997), complex both in

severity and number of symptoms as well as in duration and in frequency of remittance/recovery and recurrence, plus complex in combinations of all these elements. It might be beneficial for researchers to clarify the separate dimensions of depression as 1) severity (usually number of symptoms); 2) duration of those symptoms at that level of severity; and 3) longitudinal pattern of severity and its duration. It also would be helpful to consider all of these dimensions on a continuum rather than in terms of discrete categories. Severity, for example, conceptualized on a continuum would not require thresholds and could include a severity of level of zero symptoms or non-depressed, for example, as in recovery. It might be premature at this time to label with a diagnosis a particular combination of severity, duration, and longitudinal pattern until the comprehensive spectrum of the phenomenon of depressive symptoms is better identified than it currently is.

The theory that depressions are caused by or somehow related to life events and stressors gets muddled when there do not seem to be commonality of such events between women or even within the individual woman. It might be useful in investigating models of depression to ask, if some events are depressing to some women at some times, what is it about that event(s) at that time(s) and that woman that led to depression? What factors might be causal for each particular longitudinal pattern are still mostly unknown and worthy of further investigation. A model might need to be developed that takes into account more than a listing of life events without context.

### Clinical Implications

The clinical implications of these patterns can be included as part of midlife women's health care. First, the large group of women stable without symptoms has been easily overlooked and ignored. Both midlife women and their health care providers can be assured or reassured that those women who started midlife without symptoms were not likely to worsen, and that this was the single largest group of women.

Secondly, while the majority of women showed no trend over time, there was a modest number of women who tended to improve in symptoms over the years. Improvement with age should again be reassuring. It may be important for health care providers to consider, however, that these women, as well as the women with worsening symptoms (rising or positive slopes) often demonstrated a modest to large variability over the years. Clinically, it might be useful to understand that improvement in depressive symptoms at one time point might not be stable, and that more vigilance would be warranted for women who have had high levels of symptoms in the past. Finally, it behooves clinicians to remember that those women with rising or worsening depressive symptoms did not demonstrate any consistent relationship between changing CES-D scores and MTS.

Nevertheless, some women became depressed during the menopausal transition (as in the Blip Up pattern), and for some of them it might be a first time in years or in a lifetime (Avis et al., 1994; Winokur, 1973). New depression in midlife might be explained, however, by stable incidence rates rather than MTS (Hällström & Samuelsson, 1985). Incidence rate is "the number of disease onsets in the population" per time period

(Rothman, 1986 p. 24), in other words, the first lifetime occurrence of a health problem. If incidence rates were stable over the adult lifetime, there would be as many first-time depressions per 1000 never-depressed women in midlife as there would be first-time depressions per 1000 never-depressed women in their twenties and in their early thirties. Consistent with this hypothesis, Chen et al., (2000) found the mean age of first lifetime depressive episode as 32 years for moderate and severe episodes and 34 years for mild. An assumption that a first-time depressive episode in midlife is caused by menopausal transition does not seem to be well-founded (see Chapter 2).

Anticipation of recurrence, identified here as a seesaw pattern or high variability around a stable or rising level, might be appropriate in nursing care. Women with a history of depressive symptoms might be helped by both fairly frequent monitoring and anticipatory guidance. In a grounded theory investigation of women who had recovered from major depression, Schreiber (1996) found that all fully recovered women “knew that it could happen to them again” (p. 172). Schreiber (1996) was one of the few researchers who suggested a nursing intervention of anticipatory guidance for women who “believe they will never see depression again” (p. 172), women she described as not yet fully recovered. Given that variability over the years can be large and not uncommon, treatment modalities that have long-term effects and self-management components, such as cognitive behavioral therapy, might be most appropriate (Evans et al., 1992; Frank et al, 2000; Kupfer et al., 1992; Miller & Keitner, 1996; Shea et al., 1992). Nurses can provide interventions for depressed midlife women appropriate for the

chronic and recurring nature of this mental health problem, and with a focus appropriately on depression instead of on menopause.

The idea that a woman has her “own” level, with no trend, which was identified in the majority of women, and has been suggested as related to stable characteristics of the individual (Duncan-Jones et al., 1990; Ormel & Wohlfarth, 1991), makes the adoption of recent treatment approaches unsatisfactory. In order to cut payor costs, a “quick fix” has been advocated and even required. In this scenario the primary health care provider’s office visit includes a questionnaire (Hamilton or Beck) validated by a few questions from the primary care provider. A diagnosis of major depressive disorder is treated with no more than a prescription for an antidepressant. This approach has already been criticized as ineffective (Frank et al. 2000; Miller & Keitner, 1996; Pajer, 1995; Sturm & Wells, 1995). Further, the idea, related to the “quick fix,” that depression should be treated as an isolated episode in a woman’s life ignores the chronic nature of the problem. Health care providers and payors could provide better services if they recognized depression as a chronic diagnosis with episodes, remissions, recoveries, recurrences and relapses (APA, 1994; First et al., 1996; McCullough et al., 1996).

Other interventions have been seen in the social marketing of anti-depressant medications and self-recognition of symptoms via media advertisements. In addition there have been campaigns of professional education to increase primary care practitioners’ awareness of depression, its presentation (especially somatic), and the benefits of diagnosis and treatment (Betrus, Elmore, & Hamilton, 1995; Bridges & Goldberg, 1985; Depression Guideline Panel, 1993; Goldberg, 1979; Kirmayer &

Robbins, 1996; Kirmayer, Robbins, Dworkind, & Yaffe, 1993; Worsley, Walters, & Wood, 1977).

Primary prevention probably has to start early in life, considering the typical first episode starts before age 30 years (Chen et al. 2000; Keller et al., 1995), and for many before age 20 years (Angold, Weissman, Wickramaratne, & Prusoff, 1991; Birmaher, Ryan, Williamson, Brent, & Kaufman, 1996; Gjerde et al., 1988; Hankin et al., 1998; Lewinsohn, Clarke, Seeley, & Rohde, 1994; Lewinsohn, Rohde, Klein, & Seeley, 1999; Nolen-Hoeksema, 1990, Chap. 2; Poznanski, 1982; Rohde, Lewinsohn, & Seeley, 1994). Specifically, primary prevention must include prevention of childhood abuse and adult harassment and assault.

#### Limitations

The CES-D is no longer considered the state of the art in measurement of depression (Kohn, Dohrenwend, & Mirotznik, 1998). Several studies investigated the cut-off total CES-D score of 16 in different samples (see Chapter 2). Furukawa et al. (1997) found that receiver operating characteristic (ROC) curve estimation and/or likelihood ratio estimation provided different cut-offs in clinical and community samples. Various structured interview schedules that can be administered by trained lay personnel, such as the Diagnostic Interview Schedule (Anthony, Folstein, Romanoski, Von Korff, Nestadt, Chahal et al., 1985; Helzer, Robins, McEvoy, Spitznagel, Stoltzman, Farmer et al., 1985; Robins, Helzer, Croughan, & Ratcliff, 1981) are measures that now are considered to be more valid and reliable in identifying MDD (Kohn et al., 1998). The gold standard is considered expert diagnosis by an experienced psychiatric professional

(Chen et al., 2000; Keller et al., 1995). However, research designed and conducted in the 1980s did not have the benefit of this latest knowledge. The SMWHS original grant was written in 1988 before critiques and criticisms of the CES-D were published.

Longitudinal studies utilizing repeated measures are limited to the best available tools at the time they start and cannot later switch measures.

Measurement of the CES-D once a year limited the investigation of changes in depressive symptoms throughout the year that might have been occurring between measurement time points. Studies of duration and recurrence indicate that, especially with treatment, a three- or four-month time frame might be required to fully explore longitudinal patterns of depressive symptoms (Evans et al., 1992; Kupfer et al., 1992; Lavori, 1994; Shea et al., 1992). It is possible that if the CES-D had been administered every six or perhaps even every three months more complex patterns would have been identified. Since the proposal of RBD (Angst & Hochstrasser 1994), attention to daily changes in depressive symptoms, such as SMWHS has available in daily diaries, might elucidate the minimum frequency and period of waveforms.

The parent study contained no measure of personality type that provided a stable score across years. While the SCL-90 of Derogatis (Derogatis et al., 1976) was administered once every three or four years, the correlations between any two administrations were so low as to be unusable as a measure of a stable trait such as personality.

Another feature of the parent study was administration of the CES-D irrespective of phase of menstrual cycle or day of the week. Fluctuations in depressive symptoms

might have occurred by chance because of differences related to menstrual phase at time of completion from one year to the next, assuming a relationship between menstrual cycle phase and depressive symptoms. While last menstrual period date (LMP) was also part of the same questionnaire containing the CES-D and menstrual phase of questionnaire date could presumably be calculated from the LMP, there was no certainty that the date written on the questionnaire was the date the CES-D was completed. In fact, it was likely that there were several days difference between completion of the CES-D and the date, particularly in later years when the annual questionnaire became very long, thus putting date of CES-D completion in another menstrual cycle phase entirely. Some of the same uncertainty about exact date of CES-D completion applied to day of the week, which might have an effect on depressive symptoms.

Effects of seasonality, such as might be demonstrated in SAD, were not controlled, although note of change in seasonal completion of the annual questionnaire was mentioned in the phenomenological analysis where appropriate. However, timing of the annual questionnaire was converted from an anniversary date, evenly distributed throughout the year, to a winter season. Since in Seattle, prevalence of SAD is higher than for the U. S. as a whole (Betrus & Elmore, 1991) it was possible that conversion of the CES-D completion to winter for most women could have increased depressive symptoms in later years from what they would have been during summer for women with SAD. Table 4, however, demonstrates the converse; annual mean CES-D scores have been slightly decreasing as women aged, with no remarkable rise or even changes in trend occurring between CES-D 36 and CESD95.

As discussed in Chapter 2, measurements of predictors simultaneous with measurement of the CES-D are expected to be affected by current mood. In this case, current mood would function as both a confounder and as part of the outcome variable. In the current study, however, unlike in the other longitudinal studies of depressive symptoms during the menopausal transition, MTS was not determined retrospectively by interview questions about menstrual periods; in the SMWHS, MTS was determined by prospective recording on menstrual calendars, with only supplemental information provided by interview questions retrospective for one year's time. Likewise, life events and stressors mostly were determined by concurrent recording on the back of the same menstrual calendars, although those events that were hand written on annual questionnaires were, indeed, recalled from the previous 12 months. Lastly, recall of any history of sexual abuse/assault might have been influenced by concurrent mood in 1999 (see Chapter 2). It is also likely that this history was underreported in general (Walker et al., 1993). If underreporting were different in different clusters, the significance of results might have been affected.

In studying depressive symptoms longitudinally the issue of selection bias must be considered. While some women did demonstrate very high scores, a few even for many years, it was possible that generally the most severely depressed women were not able to continue in the study, or that women dropped out when they became depressed. Two t-tests were performed on the initial CES-D score from the entire SMWHS sample of 508 women. The first t-test found no significant differences in total CES-D scores between those women who dropped out after the initial year (mean score=11.95) and

those who continued to at least a second year (mean score=11.64),  $t = -.372$ ,  $df = 506$ ,  $p = .710$ . A second t-test, while not quite reaching statistical significance, showed lower initial CES-D scores in those women who provided yearly scores through 2000 without ever missing a questionnaire (mean score=10.67) compared to all others (mean score=12.13),  $t = -1.720$ ;  $df = 506$ ,  $p = .086$ . A final ANOVA was performed on the initial CES-D score between 10 groups, each group having a different number of annual questionnaires completed, one through 10. Again there were no significant differences and not even any particular trend in lower scores with increasing numbers of annual scores;  $F = 1.674$ ,  $df = 9,498$ ,  $p = .092$ . The lack of significance was reassuring that differential drop-out rates were not likely to be an important bias.

#### Implications For Further Research

Statistical methods designed for analysis of multi-point longitudinal data are becoming more available and recognized. Future plans include use of hierarchical data modeling (Bryk & Raudenbush, 1992) to further investigate the relationships between theorized predictors and depressive symptoms. For example, endocrine data were available for many of the women, starting from 1997, and could be added to a model predicting total CES-D score. Life events were measured in standard fashion with the Norbeck adaptation of the Sarason Life Events Survey (Norbeck, 1984) on four occasions during the 10 years and could also be included as a predictor of CES-D score. More frequent measurement of depressed mood and other depressive symptoms, as well as global stress ratings, are available intermittently from daily diary records for most of the

women; these measures could also be used to investigate fluctuations of amplitude, frequency, and period within longitudinal patterns using day or month as a unit of time.

Measures that are not available yet in the SMWHS could be added in the future. Addition of a measure of personality type would enhance a model that could be tested statistically. Other variables such as hospitalization for mental health problems could be included, too. Lastly, a prospective record of interventions other than prescriptive antidepressant use, such as cognitive behavioral therapy or other forms of psychotherapy would also contribute to understanding of longitudinal patterns of depressive symptoms in midlife women.

This data set was unique in having followed a community-based sample of women annually from 6 to 10 years. However, the data were an arbitrary window, during midlife, out of a lifetime of depressive symptoms or their absence. In spite of the fact that six, let alone 10, years is a very long time to repeatedly measure CES-D scores, this time frame is still a small portion of a woman's adult life. What has been presented in this analysis is an arbitrary snapshot of longitudinal patterns during midlife. Lifetime patterns might be even more complex and revealing than recent efforts to address some of these complexities in new diagnoses (First et al., 1996) and than the results of the current study have identified.

### Reference List

- Cluster. (1997). SPSS 7.5 Statistical algorithms (pp. 54-58). Chicago: SPSS Inc.
- Cluster analysis. (1997). In SPSS base 7.5 applications guide (pp. 263-285). Chicago: SPSS Inc.
- Hierarchical cluster analysis. (1997). In SPSS base 7.5 for windows user's guide, Chap. 25-26, (pp. 263-276). Chicago: SPSS, Inc.
- Research on the menopause in the 1990s. Report of a WHO scientific group. (1996). Geneva: World Health Organization.
- Research on the menopause. Report of a WHO scientific group. (1981). Geneva: World Health Organization.
- Using crosstabs to obtain crosstabulations and measures of association. (1997). In SPSS Base 7.5 Applications Guide (pp. 63-87). Chicago: SPSS Inc.
- Akiskal, H. S. (1988). Personality as a mediating variable in the pathogenesis of mood disorders: Implications for theory, research, and prevention. In T. Helgason, & R. J. Daly (Eds), Depressive illness: prediction of course and outcome (pp. 131-146). London, Berlin, New York: Springer-Verlag.
- Akiskal, H. S. (1994). Dysthymic and cyclothymic depressions: therapeutic considerations. Journal of Clinical Psychiatry, *55*(4 suppl), 46-52.
- Akiskal, H. S. (1985). Interaction of biologic and psychologic factors in the origin of depressive disorders. Acta Psychiatrica Scandinavica, *71*(Suppl 319), 131-139.
- Akiskal, H. S., Hirschfeld, R. M. A., & Yerevanian, B. I. (1983). The relationship of personality to affective disorders. Archives of General Psychiatry, *40*(7), 801-810.
- Akiskal, H. S. (1982). Factors associated with incomplete recovery in primary depressive illness. Journal of Clinical Psychiatry, *43*(7), 266-271.
- Akiskal, H. S., King, D., Rosenthal, T. L., Robinson, D., & Scott-Strauss, A. (1981). Chronic depression: Part I. Clinical and familiar characteristics in 137 probands. Journal of Affective Disorders, *3*, 297-315.
- Aldenderfer, M. S., & Blashfield, R. K. (1984). Cluster Analysis. Beverly Hills and London: Sage.
- Aldwin, C. M., Levenson, M. R., Spiro III, A., & Bosse, R. (1989). Does emotionality predict stress? Findings from the normative aging study. Journal of Personality and Social Psychology, *56*(4), 618-624.

Altman, N., Sachar, E. J., Gruen, P. H., Halpern, F. S., & Eto, S. (1975). Reduced plasma LH concentration in postmenopausal depressed women. Psychosomatic Medicine, 37(3), 274-6.

American Psychiatric Association Committee on Nomenclature and Statistics. (1994). Mood disorders. In Diagnostic and statistical manual of mental disorders (DSM-IV) (4<sup>th</sup> ed.) (pp. 317-391). Washington, DC: Author.

Amsterdam, J. D., Winokur, A., Lucki, I., & Snyder, P. (1983). Neuroendocrine regulation in depressed postmenopausal women and healthy subjects. Acta Psychiatrica Scandinavica, 67(1), 43-49.

Anderson, J. E., Nelson, D. E., & Wilson, R. W. (1998). Telephone coverage and measurement of health risk indicators: Data from the National Health Interview Survey. American Journal of Public Health, 88(9), 1392-1395.

Angold, A., Weissman, M., John, K., Wickramaratne, P., & Prusoff, B. (1991). The effects of age and sex on depression ratings in children and adolescents. Journal of the American Academy of Child and Adolescent Psychiatry, 30(1), 67-74.

Angst, J. (1988). Clinical course of affective disorders. In T. Helgason, & R. J. Daly (Eds), Depressive illness: prediction of course and outcome (pp. 1-48). London, Berlin, New York: Springer-Verlag.

Angst, J., & Hochstrasser, B. (1994). Recurrent brief depression: the Zurich study. Journal of Clinical Psychiatry, 55(4 suppl), 3-9.

Angst, J., & Merikangas, K. (1997). The depressive spectrum: diagnostic classification and course. Journal of Affective Disorders, 45(1-2), 31-40.

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. Archives of General Psychiatry, 42, 667-675.

Asher, H. B. (1983). Causal modeling. 2<sup>nd</sup> ed.). Newbury Park, California: Sage.

Avis, N. E., Brambilla, D., McKinlay, S. M., & Vass, K. (1994). A longitudinal analysis of the association between menopause and depression: Results from the Massachusetts Women's Health Study. Ann Epidemiol, 4, 214-220.

Avis, N. E., Crawford, S. L., & McKinlay, S. M. (1997). Psychosocial, behavioral, and health factors related to menopause symptomatology. Women's Health: Research on Gender, Behavior, and Policy, 3(2), 103-120.

Baker, A., Simpson, S., & Dawson, D. (1997). Sleep disruption and mood changes associated with menopause. Journal of Psychosomatic Research, 43(4), 359-69.

Ballinger, C. B. (1990). Psychiatric aspects of the menopause. British Journal of Psychiatry, 156, 773-787.

Bareford, C. G. (1990). A continued analysis of women's health in middle age. Annals of the New York Academy of Sciences, 592, 412-413.

Beekman, A. T., Deeg, D. J., Van Limbeek, J., Braam, A. W., De Vries, M. Z., & Van Tilburg, W. (1997). Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community-based sample of older subjects in The Netherlands. Psychological Medicine, 27(1), 231-5.

Bernard, J. S. (1981). The female world. New York: Free Press.

Betrus, P. A., & Elmore, S. K. (1991). Seasonal affective disorder, Part I: a review of the neural mechanisms for psychosocial nurses. Archives of Psychiatric Nursing, V(6), 357-364.

Betrus, P. A., Elmore, S. K., & Hamilton, P. A. (1995). Women and somatization: unrecognized depression. Health Care for Women International, 16 (4 Jul-Aug), 287-97.

Beyene, Y. (1986). Cultural significance and physiological manifestations of menopause: a biocultural analysis. Culture, Medicine, and Psychiatry, 10(1), 47-71.

Birmaher, B., Ryan, N. D., Williamson, D. E., Brent, D. A., & Kaufman, J. (1996). Childhood and adolescent depression: a review of the past 10 years. Part II. Journal of the American Academy of Child and Adolescent Psychiatry, 35(12), 1575-83.

Bland, R. C. (1997). Epidemiology of affective disorders: A review. Canadian Journal of Psychiatry, Revue Canadienne De Psychiatrie, 42(4), 367-77.

Bolger, N., & Schilling, E. A. (1991). Personality and the problems of everyday life: The role of neuroticism in exposure and reactivity to daily stressors. Journal of Personality, 59(3), 355-386.

Boyd, J. H., Weissman, M. M., Thompson, D., & Myers, J. K. (1982). Screening for depression in a community sample. Archives of General Psychiatry, 39(Oct), 1195-1200.

Boyle, G. J. (1985). Self-report measures of depression: Some psychometric considerations. British Journal of Clinical Psychology, 24(Pt 1), 45-59.

Bracke, P. (1998). Sex differences in the course of depression: evidence from a longitudinal study of a representative sample of the Belgian population. Social Psychiatry and Psychiatric Epidemiology, 33(9), 420-9.

Breslau, N. (1985). Depressive symptoms, major depression, and generalized anxiety: a comparison of self-reports on CES-D and results from diagnostic interviews. Psychiatry Research, 15, 219-229.

Breslau, N., Schultz, L., & Peterson, E. (1995). Sex differences in depression: a role for preexisting anxiety. Psychiatry Research, 58, 1-12.

Bridges, K. W., & Goldberg, D. P. (1985). Somatic presentation of DSM III psychiatric disorders in primary care. Journal of Psychosomatic Research, 29(6), 563-569.

Brieger, P., & Marneros, A. (1997). Dysthymia and cyclothymia: historical origins and contemporary development. Journal of Affective Disorders, 45(1-2), 117-126.

Briscoe, M. (1982). Sex differences in psychological well-being. Psychological Medicine, Monograph Supplement 1.

Broadhead, W. E., Blazer, D. G., & George, L. K. (1990). Depression, disability days, and days lost from work in a prospective epidemiologic survey. JAMA, 264(19), 2524-8.

Bromberger, J. T., & Matthews, K. A. (1996a). A "feminine" model of vulnerability to depressive symptoms: A longitudinal investigation of middle-aged women. Journal of Personality and Social Psychology, 70(3), 591-598.

Bromberger, J. T., & Matthews, K. A. (1996b). A longitudinal study of the effects of pessimism, trait anxiety, and life stress on depressive symptoms in middle-aged women. Psychology and Aging, 11(2), 207-213.

Brown, G. W. (1998). Loss and depressive disorders. In Adversity, stress, and psychopathology (pp. 358-370). New York, Oxford: Oxford University Press.

Brown, G. W., & Harris, T. (1978). Social origins of depression: a study of psychiatric disorder. New York: Free Press.

Bryk, A. S., & Raudenbush, S. W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, California: Sage.

Burger, H. G., Dudley, E. C., Hopper, J. L., Shelley, J. M., Green, A., Smith, A., Dennerstein, L., & Morse, C. (1995). The endocrinology of the menopausal transition: a

cross-sectional study of a population-based sample. Journal of Clinical Endocrinology and Metabolism, 80(12), 3537-45.

Burt, V. K., Altshuler, L. L., & Rasgon, N. (1998). Depressive symptoms in the perimenopause: Prevalence, assessment, and guidelines for treatment. Harvard Review of Psychiatry, 6(3), 121-132.

Busch, C. M., Zonderman, A. B., & Costa Jr., P. T. (1994). Menopausal transition and psychological distress in a nationally representative sample: is menopause associated with psychological distress? Journal of Aging and Health, 6(2), 209-228.

Byrne, B. M. (1989). A primer of LISREL: Basic applications and programming for confirmatory factor analytic models. New York: Springer-Verlag.

Callahan, C. M., Hui, S. L., Nienaber, n. A., Musick, B. S., & Tierney, W. M. (1994). Longitudinal study of depression and health services use among elderly primary care patients. Journal of the American Geriatrics Society, 42(8), 833-8.

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. Psychological Review, 97(1), 19-35.

Chen, L., Eaton, W. W., Gallo, J. J., & Nestadt, G. (2000). Understanding the heterogeneity of depression through the triad of symptoms, course and risk factors: a longitudinal, population-based study. Journal of Affective Disorders, 59(1), 1-11.

Chiazze Jr., L., Brayer, F. T., Macisco Jr., J. J., Parker, M. P., & Duffy, B. J. (1968). The length and variability of the human menstrual cycle. JAMA, 203(6 (Feb 5)), 377-380.

Cho, M. J., Moscicki, E. K., Narrow, W. E., Rae, D. S., Locke, B. Z., & Regier, D. A. (1993). Concordance between two measures of depression in the Hispanic Health and Nutrition Examination Survey. Social Psychiatry and Psychiatric Epidemiology, 28, 156-163.

Clark, V. A., Aneshensel, C. S., Frerichs, R. R., & Morgan, T. M. (1981). Analysis of effects of sex and age in response to items on the CES-D scale. Psychiatry Research, 5, 171-181.

Cole, S. R., Kawachi, I., Maller, S. J., & Berkman, L. F. (2000). Test of item-response bias in the CES-D scale. experience from the New Haven EPESE study. Journal of Clinical Epidemiology, 53(3), 285-9.

Coleman, P. M. (1993). Depression during the female climacteric period. Journal of Advanced Nursing, 18(10), 1540-6.

Collins, A., & Landgren, B.-M. (1997). Psychosocial factors associated with the use of hormonal replacement therapy in the longitudinal follow-up of Swedish women. Maturitas, 28, 1-9.

Collins, L. M., & Horn, J. L. (Eds). (1991). Best methods for the analysis of change: recent advances, unanswered questions, future directions. Washington, D. C.: American Psychological Association.

Comstock, G. W., & Helsing, K. J. (1976). Symptoms of depression in two communities. Psychological Medicine, 6, 551-563.

Cornoni-Huntley, J., Barbano, H. E., Brody, J. A., Cohen, B., Feldman, J. J., Kleinman, J. C., & Madans, J. (1983). National Health and Nutrition Examination I--epidemiologic follow-up survey. Public Health Reports, 98(3), 245-251.

Coryell, W. (1997). Do psychotic, minor and intermittent depressive disorders exist on a continuum? Journal of Affective Disorders, 45(1-2), 75-83.

Coryell, W., Endicott, J., & Keller, M. B. (1991). Predictors of relapse into major depressive disorder in a nonclinical population. American Journal of Psychiatry, 148(10), 1353-1358.

Coryell, W., Scheftner, W., Keller, M., Endicott, J., Maser, J., & Klerman, G. L. (1993). The enduring psychosocial consequences of mania and depression. American Journal of Psychiatry, 150(5), 720-727.

Costa Jr., P. T., Zonderman, A. B., McCrae, R. R., Cornoni-Huntley, J., Locke, B. Z., & Barbano, H. E. (1987). Longitudinal analyses of psychological well-being in a national sample: stability of mean levels. Journal of Gerontology, 42(1), 50-55.

Cowdry, R. W., Gardner, D. L., O'Leary, K. M., Leibenluft, E., & Rubinow, D. R. (1991). Mood variability: a study of four groups. American Journal of Psychiatry, 148(11), 1505-11.

Craig, T. J., & Van Natta, P. A. (1979). Influence of demographic characteristics on two measures of depressive symptoms. Archives of General Psychiatry, 36(February), 149-154.

Craig, T. J., & Van Natta, P. A. (1976). Presence and persistence of depressive symptoms in patient and community populations. American Journal of Psychiatry, 133(12), 1426-9.

Crawford, G. (1982). The concept of pattern in nursing: conceptual development and measurement. ANS [Advances in Nursing Science], 5(1), 1-6.

Davis, D. L. (1986). The meaning of menopause in a Newfoundland fishing village. Culture, Medicine, and Psychiatry, 10(1), 73-94.

Dennerstein, L., Dudley, E., & Burger, H. (1997). Well-being and the menopausal transition. Journal of Psychosomatic Obstetrics and Gynaecology, 18(2), 95-101.

Dennerstein, L., Smith, A. M. A., & Morse, C. (1994). Psychological well-being, mid-life and the menopause. Maturitas, 20(1), 1-11.

Dennerstein, L., Smith, A. M. A., Morse, C., Burger, H., Green, A., Hopper, J., & Ryan, M. (1993). Menopausal symptoms in Australian women. Medical Journal of Australia, 159, 232-236.

Depression Guideline Panel. (1993). Depression in Primary Care: Volume 1. detection and diagnosis (AHCPR Publication No. 93-0550). Rockville, Maryland: U. S. Department of Health and Human Services.

Derby, C. A., Hume, A. L., Barbour, M. M., McPhillips, J. B., Lasater, T. M., & Carleton, R. A. (1993). Correlates of postmenopausal estrogen use and trends through the 1980s in two southeastern New England communities. American Journal of Epidemiology, 137(10), 1125-1135.

Derogatis, L. R., Rickels, K., & Rock, A. F. (1976). The SCL-90 and the MMPI: a step in the validation of a new self-report scale. British Journal of Psychiatry, 128, 280-289.

Detre, T. (1968). Severe emotional reactions precipitated by the climacterium. Psychosomatics, 9(4), Suppl:31-5.

Devins, G. M., & Orme Carolee M. (1985). Center for Epidemiological Studies Depression scale. In Keyser D. J., & R. C. Sweeland (Eds), Test critiques Vol. 2 (pp. 144-160). Kansas City, Mo: Test Corporation of America.

Dohrenwend, B. P. (1998). Overview of evidence for the importance of adverse environmental conditions in causing psychiatric disorders. In Adversity, stress, and psychopathology (pp. 523-538). New York, Oxford: Oxford University Press.

Duncan-Jones, P., Fergusson, D. M., Ormel, J., & Horwood, L. J. (1990). A model of stability and change in minor psychiatric symptoms: results from three longitudinal studies. Psychological Medicine, Monograph Supplement 18.

Ekstrom, D., Quade, D., & Golden, R. N. (1990). Statistical analysis of repeated measures in psychiatric research. Archives of General Psychiatry, 47, 770-772.

Elmore, S. K., Betrus, P. A., & Burr, R. (1994). Light, social zeitgebers, and the sleep-wake cycle in the entrainment of human circadian rhythms. Research in Nursing & Health, 17(6), 471-478.

Ensel, W. M., & Lin, N. (1991). The life stress paradigm and psychological distress. Journal of Health & Social Behavior, 32(4), 321-341.

Evans, M. D., Hollon, S. D., DeRubeis, R. J., Piasecki, J. M., Grove, W. M., Garvey, M. J., & Tuason, V. B. (1992). Differential relapse following cognitive therapy and pharmacotherapy for depression. Archives of General Psychiatry, 49(10), 802-808.

Faulstich, M. E., Carey, M. P., Ruggiero, L., Enyart, P., & Gresham, F. (1986). Assessment of depression in childhood and adolescence: an evaluation of the Center for Epidemiological Studies Depression scale for children (CES-DC). American Journal of Psychiatry, 143(8), 1024-1027.

Fergusson, D. M., & Horwood, L. J. (1987). Vulnerability to life events exposure. Psychological Medicine, 17 (3), 739-749.

First, M. B., Donovan, S., & Frances, A. (1996). Nosology of chronic mood disorders. Psychiatric Clinics of North America, 19(1), 29-39.

Fischer, C. T. (1975). Intelligence contra IQ: A human science critique and alternative to the natural science approach to man. In A. Giorgi, C. T. Fisher, & E. L. Murray (Eds), Duquesne studies in phenomenological psychology Vol. 2, Chap. 9, (pp. 143-154). Pittsburgh, PA: Duquesne University Press.

Fisher, L. D., & Belle, G. V. (1993). Biostatistics: A methodology for the health sciences. New York: John Wiley & Sons, Inc.

Fleiss, J. L. (1981). Statistical methods for rates and proportions. (2nd ed.). New York: John Wiley & Sons.

Foelker, G. A., & Shewchuk, R. M. (1992). Somatic complaints and the CES-D. Journal of the American Geriatrics Society, 40(3), 259-62.

Fogel, C. I., & Woods, N. F. (1995). Midlife women's health. In C. I. Fogel, & N. F. Woods (Eds), Women's Health Care: a comprehensive handbook (pp. 79-100). Thousand Oaks, Calif.: Sage.

Frank, E., Grochocinski, V. J., Spanier, C. A., Buysse, D. J., Cherry, C. R., Houck, P. R., Stapf, D. M., & Kupfer, D. J. (2000). Interpersonal psychotherapy and antidepressant medication: evaluation of a sequential treatment strategy in women with recurrent major depression. Journal of Clinical Psychiatry, 61(1), 51-57.

- Frerichs, R. R., Aneshensel, C. S., & Clark, V. A. (1981). Prevalence of depression in Los Angeles County. American Journal of Epidemiology, 113(6), 691-699.
- Friedman, R. A., & Kocsis, J. H. (1996). Pharmacotherapy for chronic depression. Psychiatric Clinics of North America, 19(1), 121-132.
- Furukawa, T., Hirai, T., Kitamura, T., & Takahashi, K. (1997). Application of the Center for Epidemiologic Studies Depression Scale among first-visit psychiatric patients: a new approach to improve its performance. Journal of Affective Disorders, 46, 1-13.
- Furukawa, T. A., Konno, W., Morinobu, S., Harai, H., Kitamura, T., & Takahashi, K. (2000). Course and outcome of depressive episodes: comparison between bipolar, unipolar and subthreshold depression. Psychiatry Res, 96(3), 211-20.
- Garrison, C. Z., Addy, C. L., Jackson, K. L., McKeown, R. E., & Waller, J. L. (1991). The CES-D as a screen for depression and other psychiatric disorders in adolescents. Journal of the American Academy of Child and Adolescent Psychiatry, 30(4), 636-641.
- Gath, D., Cooper, P., & Day, A. (1982). Hysterectomy and psychiatric disorder: I. Levels of psychiatric morbidity before and after hysterectomy. British Journal of Psychiatry, 140, 335-342.
- Gath, D., & Iles, S. (1990). Depression and the menopause. BMJ, 300 May 19(6735), 1287-8.
- Gatz, M., & Hurwicz, M.-L. (1990). Are old people more depressed? Cross-sectional data on Center for Epidemiological Studies Depression Scale factors. Psychology and Aging, 5(2), 284-90.
- Gaynes, B. N., Magruder, K. M., Burns, B. J., Wagner, H. R., Yarnall, K. S. H., & Broadhead, W. E. (1999). Does a coexisting anxiety disorder predict persistence of depressive illness in primary care patients with major depression? General Hospital Psychiatry, 21, 158-167.
- George, T. (1988). Menopause: some interpretations of the results of a study among a non-western group. Maturitas, 10, 109-116.
- Gerety, M. B., Williams Jr., J. W., Mulrow, C. D., Cornell, J. E., Kadri, A. A., Rosenberg, J., Chiodo, L. K., & Long, M. (1994). Performance of case-finding tools for depression in the nursing home: influence of clinical and functional characteristics and selection of optimal threshold scores. Journal of the American Geriatrics Society, 42, 1103-1109.

Gershensfeld, N. A., & Weigend, A. S. (1997). The future of time series: Learning and understanding. In H. F. Nijhout, L. Nadel, & D. Stein (Eds), Pattern formation in the physical and biological sciences (pp. 349-429). Reading, MA: Addison-Wesley.

Giancoli, D. C. (1998). Physics: Principles with applications. (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Giel, R. (1998). Natural and human-made disasters. In B. P. Dohrenwend (Ed), Adversity, stress, and psychopathology (pp. 66-79). New York, Oxford: Oxford University Press.

Giles, D. E., Jarret, R. B., Biggs, M. M., Guzick, D. S., & Rush, A. J. (1989). Clinical predictors of recurrence in depression. American Journal of Psychiatry, 146(6), 764-767.

Giorgi, A. (1975). Convergence and divergence of qualitative and quantitative methods in psychology. In A. Giorgi, C. T. Fisher, & E. L. Murray (Eds), Duquesne studies in phenomenological psychology Vol. 2, Chap. 5, (pp. 72-9). Pittsburgh, PA: Duquesne University Press.

Gjerde, P. F., Block, J., & Block, J. H. (1988). Depressive symptoms and personality during late adolescence: gender differences in the externalization-internalization of symptom expression. Journal of Abnormal Psychology, 97(4), 475-486.

Goldberg, D. (1979). Detection and assessment of emotional disorders in a primary-care setting. International Journal of Mental Health, 8(2), 3048.

Goldberg, R. T. (1994). Childhood abuse, depression, and chronic pain. Clinical Journal of Pain, 10(4), 277-81.

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, 53(4), 461-469.

Gotlib, I. H., Lewinsohn, P. M., & Seeley, J. R. (1995). Symptoms versus a diagnosis of depression: differences in psychosocial functioning. Journal of Consulting and Clinical Psychology, 63(1), 90-100.

Greene, J. G. (1992). The cross-sectional legacy: an introduction to longitudinal studies of the climacteric. Maturitas, 14, 95-101.

Gruen, P. H., Sachar, E. J., Altman, N., & Sassin, J. (1975). Growth hormone responses to hypoglycemia in postmenopausal depressed women. Archives General Psychiatry, 32(1), 31-3.

Guarnaccia, P. J., Angel, R., & Worobey, J. L. (1989). The factor structure of the CES-D in the Hispanic health and nutrition examination survey: the influences of ethnicity, gender and language. *Social Science & Medicine*, 29(1), 85-94.

Guicheney, P., Leger, D., Barrat, J., Trevoux, R., De Lignieres, B., Roques, P., Garnier, J. P., Boyer, P., Grenier, J., Dreux, C., & et, a. l. (1988). Platelet serotonin content and plasma tryptophan in peri- and postmenopausal women: variations with plasma oestrogen levels and depressive symptoms. *European Journal of Clinical Investigation*, 18(3), 297-304.

Gwirtsman, H. E., Blehar, M. C., McCullough, J. P.Jr., Kocsis, J. H., & Prien, R. F. (1997). Standardized assessment of dysthymia: report of a National Institute of Mental Health Conference. *Psychopharmacology Bulletin*, 33(1), 3-11.

Haan, M., Kaplan, G. A., & Camacho, T. (1987). Poverty and health: Prospective evidence from the Alameda County study. *American Journal of Epidemiology*, 125(6), 989-998.

Hall, L. A., Sachs, B., Rayens, M. K., & Luttenbacher, M. (1993). Childhood physical and sexual abuse: their relationship with depressive symptoms in adulthood. *Image: Journal of Nursing Scholarship*, 25(4), 317-23.

Hällström, T., & Samuelsson, S. (1985). Mental health in the climacteric: the longitudinal study of women in Gothenburg. *Acta Obstetrica Et Gynecologica Scandinavica, Supplement 130*, 13-18.

Hankin, B. L., Abramson, L. Y., Moffitt, T. E., Silva, P. A., McGee, R., & Angell, K. E. (1998). Development of depression from preadolescence to young adulthood: emerging gender differences in a 10-year longitudinal study. *Journal of Abnormal Psychology*, 107(1), 128-40.

Harlow, S. D., Goldberg, E. L., & Comstock, G. W. (1991). A longitudinal study of risk factors for depressive symptomatology in elderly widowed and married women. *American Journal of Epidemiology*, 134(5), 526-38.

Hays, J. C., Blazer, D. G., & Gold, D. T. (1993). CES-D: Cutpoint or change score? *Journal of the American Geriatrics Society*, 41(3), 344-345.

Heck, R. H., & Thomas Scott L. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Lawrence Erlbaum Associates.

Helzer, J. E., Robins, L. N., McEvoy Larry T., Spitznagel, E. L., Stoltzman, R. K., Farmer, A., & Brockington, I. F. (1985). A comparison of clinical and Diagnostic Interview Schedule diagnoses: physician reexamination of lay-interviewed cases in the general population. *Archives of General Psychiatry*, 42(7), 657-666.

Hemminki, E., Kennedy, D. L., Baum, C., & McKinlay, S. M. (1988). Prescribing of noncontraceptive estrogens and progestins in the United States, 1974-86. American Journal of Public Health, 78(11), 1479-1481.

Henderson, A. S. (1983). Vulnerability to depression: the lack of social support does not cause depression. In J. Angst (ed), The origins of depression: current concepts and approaches Life Sciences Research Report 26 (pp. 107-119). Berlin, Heidelberg, New York, Tokyo: Springer-Verlag.

Henderson, A. S. (1998). Social support: its present significance for psychiatric epidemiology. In Adversity, stress, and psychopathology (pp. 390-397). New York, Oxford: Oxford University Press.

Hirschfeld, R. M., Klerman, G. L., Clayton, P. J., & Keller, M. B. (1983). Personality and depression: empirical findings. Archives of General Psychiatry, 40, 993-998.

Hirschfeld, R. M. A., & Cross, C. K. (1982). Epidemiology of affective disorders. Archives of General Psychiatry, 39(1), 35-46.

Holte, A. (1992). Influences of natural menopause on health complaints: A prospective study of healthy Norwegian women. Maturitas, 14, 127-141.

Horwath, E., Johnson, J., Klerman, G. L., & Weissman, M. M. (1992). Depressive symptoms as relative and attributable risk factors for first-onset major depression. Archives General Psychiatry, 49(10), 817-823.

Huba, G. J., Lawlor, W. G., Stallone, F., & Fieve, R. R. (1976). The use of autocorrelation analysis in the longitudinal study of mood patterns in depressed patients. British Journal of Psychiatry, 128, 146-55.

Hunter, M. (1992). The south-east England longitudinal study of the climacteric and postmenopause. Maturitas, 14, 117-126.

Hunter, M., Battersby, R., & Whitehead, M. (1986). Relationships between psychological symptoms, somatic complaints and menopausal status. Maturitas, 8, 217-228.

Hunter, M. S. (1993). Predictors of menopausal symptoms: psychosocial aspects. Baillière's Clinical Endocrinology and Metabolism, 7(1), 33-45.

Hunter, M. S. (1990). Somatic experience of the menopause: A prospective study. Psychosomatic Medicine, 52(3), 357-367.

Jenkins, R. (1985). Sex differences in minor psychiatric morbidity. Psychological Medicine, Monograph Supplement 7.

Johnson, J. G., Cohen, P., Brown, J., Smailes, E. M., & Bernstein, D. P. (1999). Childhood maltreatment increases risk for personality disorders during early adulthood. Archives of General Psychiatry, 56(7), 600-606.

Johnson, J., Weissman, M. M., & Klerman, G. L. (1992). Service utilization and social morbidity associated with depressive symptoms in the community. JAMA, 267, 1478-1483.

Judd, L. L. (1997b). Pleomorphic expressions of unipolar depressive disease: summary of the 1996 CINP President's Workshop. Journal of Affective Disorders, 45(1-2), 109-116.

Judd, L. L. (1997a). Pleomorphic expressions of unipolar depressive disease: toward a new diagnostic paradigm. Journal of Affective Disorders, 45(1-2), 1-3.

Judd, L. L., Akiskal, H. S., & Paulus, M. P. (1997). The role and clinical significance of subsyndromal depressive symptoms (SSD) in unipolar major depressive disorder. Journal of Affective Disorders, 45(1-2), 5-18.

Judd, L. L., Rapaport, M. H., Paulus, M. P., & Brown, J. L. (1994). Subsyndromal symptomatic depression: a new mood disorder? Journal of Clinical Psychiatry, 55(4, suppl), 18-28.

Kaiser, K. (1990). Cross-cultural perspectives on menopause. Annals of the New York Academy of Sciences, 592, 430-432.

Kandel, D. B., Davies, M., & Raveis, V. H. (1985). The stressfulness of daily social roles for women: marital, occupational and household roles. Journal of Health & Social Behavior, 26(March), 64-78.

Katschnig, H., Angst, J., Clayton, P. J., Gershon, E. S., Hautzinger, M., Helmchen, H., Henderson, A. S., Klerman, G. L., Möller, H.-J., Paykel, E. S., Roth, M., & Weissman, M. M. (1983). Risk factors: Group report. In J. Angst (Ed), The origins of depression: current concepts and approaches Life Sciences Research Report 26 (pp. 385-403). Berlin, Heidelberg, New York, Tokyo: Springer-Verlag.

Kaufert, P., Lock, M., McKinlay, S., Beyenne, Y., Coope, J., David, D., Eliasson, M., Gonalons-Nicolet, M., Goodman, M., & Holte, A. (1986). Menopause research: the Korpilampi workshop. Social Science & Medicine, 22(11), 1285-1289.

Kaufert, P., & Syrotuik, J. (1981). Symptom reporting at the menopause. Social Science & Medicine, 15E, 173-184.

Kaufert, P. A. (1984). Women and their health in the middle years: a Manitoba project. *Social Science & Medicine*, 18(3), 279-281.

Kaufert, P. A., Gilbert, P., & Tate, R. (1987). Defining menopausal status: the impact of longitudinal data. *Maturitas*, 9, 217-226.

Kaufert, P. A., Gilbert, P., & Tate, R. (1992). The Manitoba Project: a re-examination of the link between menopause and depression. *Maturitas*, 14, 143-155.

Kay, M., Voda, A., Olivas, G., Rios, F., & Imle, M. (1982). Ethnography of the menopause-related hot flash. *Maturitas*, 4, 217-227.

Kazdin, A. E. (1982). *Single-case research designs: methods for clinical and applied settings*. New York: Oxford University.

Keitner, G. I., Ryan, C., Miller, I. W., Kohn, R., & Epstein, N. B. (1991). 12-Month outcome of patients with major depression and comorbid psychiatric or medical illness (compound depression). *American Journal of Psychiatry*, 148(3), 345-350.

Keller, M. B., Hanks, D. L., & Klein, D. N. (1996). Summary of the DSM-IV mood disorders field trial and issue overview. *Psychiatric Clinics of North America*, 19(1), 1-28.

Keller, M. B., Hirschfeld, R. M. A., & Hanks, D. (1997). Double depression: a distinctive subtype of unipolar depression. *Journal of Affective Disorders*, 45(1-2), 65-73.

Keller, M. B., Klein, D. N., Hirschfeld, R. M. A., Kocsis, J. H., McCullough, J. P., Miller, I., First, M. B., Holzer III, C. P., Keitner, G. I., Marin, D. B., & Shea, T. (1995). Results of the DSM-IV mood disorders field trial. *American Journal of Psychiatry*, 152(6), 843-849.

Keller, M. B., Lavori, P. W., Lewis, C. E., & Klerman, G. L. (1983). Predictors of relapse in major depressive disorder. *JAMA*, 250(24), 3299-3304.

Keller, M. B., Lavori, P. W., Mueller, T. I., Endicott, J., Coryell, W., Hirschfeld, R. M. A., & Shea, T. (1992). Time to recovery, chronicity, and levels of psychopathology in major depression. *Archives General Psychiatry*, 49(10), 809-816.

Keller, M. B., Lavori, P., Rice, J., Coryell, W., & Hirschfeld, R. M. A. (1986). The persistent risk of chronicity in recurrent episodes of non-bipolar major depressive disorder: a prospective follow-up. *American Journal of Psychiatry*, 143, 24.

Kessler, R. C., Borges, G., & Walters, E. E. (1999). Prevalence of and risk factors for lifetime suicide attempts in the National Comorbidity Survey. *Archives of General Psychiatry*, 56(7), 617-626.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H.-U., & Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. Archives of General Psychiatry, *51*(1), 8-19.

Kessler, R. C., Zhao, S., Blazer, D. G., & Swartz, M. (1997). Prevalence, correlates, and course of minor depression and major depression in the national comorbidity survey. Journal of Affective Disorders, *45*(1-2), 19-30.

Kilpatrick, D. G., Resnick, H. S., Saunders, B. E., & Best, C. L. (1998). Rape, other violence against women, and posttraumatic stress disorder. In Adversity, stress, and psychopathology (pp. 161-176). New York, Oxford: Oxford University Press.

Kirmayer, L. J., & Robbins, J. M. (1996). Patients who somatize in primary care: a longitudinal study of cognitive and social characteristics. Psychological Medicine, *26*, 937-951.

Kirmayer, L. J., Robbins, J. M., Dworkind, M., & Yaffe, M. J. (1993). Somatization and the recognition of depression and anxiety in primary care. American Journal of Psychiatry, *150*(5), 734-41.

Klaiber, E. L., Broverman, D. M., Vogel, W., & Kobayashi, Y. (1979). Estrogen therapy for severe persistent depressions in women. Archives of General Psychiatry, *36*(5), 550-554.

Klein, D. N., Kocsis, J. H., McCullough, J. P., Holzer, C. E. III, Hirschfeld, R. M. A., & Keller, M. B. (1996). Symptomatology in dysthymic and major depressive disorder. Psychiatric Clinics of North America, *19*(1), 41-53.

Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). Applied regression analysis and other multivariable methods. (2<sup>nd</sup> ed.). Boston: PWS-Kent Publishing.

Kohn, R., Dohrenwend, B. P., & Mirotznik, J. (1998). Epidemiological findings on selected psychiatric disorders in the general population. In Adversity, stress, and psychopathology (pp. 235-284). New York, Oxford: Oxford University Press.

Kupfer, D. J., Frank, E., Perel, J. M., Cornes, C., Mallinger, A. G., Thase, M. E., McEachran, A. B., & Grochocinski, V. J. (1992). Five-year outcome for maintenance therapies in recurrent depression. Archives General Psychiatry, *49*(10), 769-773.

Lara, M. E., & Klein, D. N. (1999). Psychosocial processes underlying the maintenance and persistence of depression: implications for understanding chronic depression. Clin Psychol Rev, *19* (5), 553-70.

Lavori, P. (1990). ANOVA, MANOVA, my black hen. Archives of General Psychiatry, *47*, 775-778.

Lavori, P. W., Dawson, R., & Mueller, T. B. (1994). Causal estimation of time-varying treatment effects in observational studies: application to depressive disorder. Statistics in Medicine, *13*(1089-1100).

Lennon, M. C. (1987). Is menopause depression? An investigation of three perspectives. Sex Roles, *17*(1-2), 1-16.

Lepine, J.-P., Pelissolo, A., Weiller, E., Boyer, P., & Lecrubier, Y. (1995). Recurrent brief depression: Clinical and epidemiological issues. Psychopathology, *28*(Suppl 1), 86-94.

Lerner, H. G. (1987). Female depression: Self-sacrifice and self-betrayal in relationships. In R. Formanek, & A. Gurian (Eds), Women and depression: a lifespan perspective (pp. 200-221). New York: Springer.

Lewinsohn, P. M., Clarke, G. N., Seeley, J. R., & Rohde, P. (1994). Major depression in community adolescents: Age at onset, episode duration, and time to recurrence. Journal of the American Academy of Child and Adolescent Psychiatry, *33*(6), 809-18.

Lewinsohn, P. M., Fenn, D. S., Stanton, A. K., & Franklin, J. (1986). Relation of age at onset to duration of episode in unipolar depression. Journal of Psychology and Aging, *1*(1), 63-68.

Lewinsohn, P. M., Rohde, P., Klein, D., & Seeley, J. R. (1999). Natural course of adolescent major depressive disorder: I. Continuity into young adulthood. Journal of the American Academy of Child and Adolescent Psychiatry, *38*(1), 56-63.

Lewinsohn, P. M., & Teri, L. (1982). Selection of depressed and nondepressed subjects on the basis of self-report data. Journal of Consulting and Clinical Psychology, *50*(4), 590-591.

Lewinsohn, P. M., Zeiss, A. M., & Duncan, E. M. (1989). Probability of relapse after recovery from an episode of depression. Journal of Abnormal Psychology, *98*(2), 107-116.

Liang, J., Tran, T. V., Krause, N., & Markides, K. S. (1989). Generational differences in the structure of the CES-D scale in Mexican Americans. Journal of Gerontology, *44*, 110-120.

Liem, J. H., James, J. B., O'Toole, J. G., & Boudewyn Arne C. (1997). Assessing resilience in adults with histories of childhood sexual abuse. American Journal of Orthopsychiatry, 67(4), 594-606.

Lin, E. H. B., Katon, W. J., Von Korff, M., Russo, J. E., Simon, G. E., Bush, T. M., Rutter, C. M., Walker, E. A., & Ludman, E. (1998). Relapse of depression in primary care: Rate and clinical predictors. Archives of Family Medicine, 7(5), 462-4.

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Beverly Hills, London, New Delhi: Sage.

Lock, M. (1986). Ambiguities of aging: Japanese experience and perceptions of menopause. Culture, Medicine and Psychiatry, 10, 23-46.

Mackey, T., Sereika, S. M., Weissfeld, L. A., Hacker, S. S., Zende, J. F., & Heard, S. L. (1992). Factors associated with long-term depressive symptoms of sexual assault victims. Archives Psychiatr Nurs, 6, 10-25.

Magursky, V., Mesko, M., & Sokolik, L. (1975). Age at the menopause and onset of the climacteric in women of Martin District, Czechoslovakia: Statistical survey and some biological and social correlations. International Journal of Fertility, 20, 17-23.

Maier, W., Gansicke, M., & Weiffenbach, O. (1997). The relationship between major and subthreshold variants of unipolar depression. Journal of Affective Disorders, 45(1-2), 41-51.

Markianos, M., Kykouras, L., & Stefanis, C. (1996). Prolactin and TSH responses to TRH and to ECT in pre- and postmenopausal women with major depression. Biological Psychiatry, 40(5), 403-406.

Matthews, K. A., Wing, R. R., Kuller, L. H., Meilahn, E. N., & Kelsey, S. F. (1990). Influences of natural menopause on psychological characteristics and symptoms of middle-aged healthy women. Journal of Consulting and Clinical Psychology, 58(3), 345-351.

Matussek, N. (1988). Biological aspects of course and outcome in depressive illness: Needed areas of research. In T. Helgason, & R. J. Daly (Eds), Depressive illness: prediction of course and outcome (pp. 92-110). London, Berlin, New York: Springer-Verlag.

McCullough, J. P., Kornstein, S. G., McCullough, J. P., Belyea-Caldwell, S., Kaye, A. L., Roberts, C., Plybon, J. K., & Kruus, L. K. (1996). Differential diagnosis of chronic depressive disorders. Psychiatric Clinics of North America, 19(1), 55-71.

McKinlay, J. B., McKinlay, S. M., & Brambilla, D. (1987). The relative contributions of endocrine changes and social circumstances to depression in mid-aged women. Journal of Health & Social Behavior, 28(December), 345-363.

McKinlay, S. M., Brambilla, D. J., & Posner, J. G. (1992). The normal menopause transition. Maturitas, 14, 103-115.

Miller, A. M., Wilbur, J., Montgomery, A. C., & Chandler, P. (1998). Social role quality and psychological well being. AAOHN Journal, 46(8), 371-8.

Miller, I. W., & Keitner, G. I. (1996). Combined medication and psychotherapy in the treatment of chronic mood disorders. Psychiatric Clinics of North America, 19(1), 151-171.

Mitchell, E. S., Woods, N. F., & Mariella, A. (2000). Three stages of the menopausal transition from the Seattle Midlife Women's Health Study: toward a more precise definition. Menopause: Journal of the North American Menopause Society, 7(5), 334-349.

Mollica, R. F., & Poole, C. T. S. (1998). Symptoms, functioning, and health problems in a massively traumatized population: The legacy of the Cambodian tragedy. In B. P. Dohrenwend (Ed), Adversity, stress, and psychopathology (pp. 34-51). New York, Oxford: Oxford University Press.

Morse, C. A., Smith, A., Dennerstein, L., Green, A., Hopper, J., & Burger, H. (1994). The treatment seeking woman at menopause. Maturitas, 18(3), 161-173.

Moscarello, R. (1991). Posttraumatic stress disorder after sexual assault: its psychodynamics and treatment. Journal of the American Academy of Psychoanalysis, 19, 235-53.

Mueller, T. I., & Leon, A. C. (1996). Recovery, chronicity, and levels of psychopathology in major depression. Psychiatric Clinics of North America, 19(1), 85-102.

Mundt, C., Reck, C., Backenstrass, M., Kronmuller, K., & Fiedler, P. (2000). Reconfirming the role of life events for the timing of depressive episodes: a two-year prospective follow-up study. Journal of Affective Disorders, 59(1), 23-30.

Myers, J. K., & Weissman, M. M. (1980). Use of a self-report symptom scale to detect depression in a community sample. American Journal of Psychiatry, 137(9), 1081-1084.

Nathanson, C. A. (1980). Social roles and health status among women: the significance of employment. Social Science & Medicine, 14A, 463-471.

Nicol-Smith, L. (1996). Causality, menopause, and depression: a critical review of the literature. BMJ, 313(7067), 1229-32.

Nijhout, H. F. (1997). Pattern and Process. In H. F. Nijhout, L. Nadel, & D. L. Stein (Eds). Pattern formation in the physical and biological sciences (pp. 1-12). Reading, MA: Addison-Wesley.

Nolen-Hoeksema, S. (1990). Sex differences in depression. Stanford, California: Stanford University Press.

Norbeck, J. S. (1984). Modification of life event questionnaires for use with female respondents. Research in Nursing & Health, 7(1), 61-71.

Nuckolls, K. B., Cassel, J., & Kaplan, B. H. (1972). Psychosocial assets, life crisis and the prognosis of pregnancy. American Journal of Epidemiology, 95(5), 431-444.

Obermeyer, C. M. (2000). Menopause across cultures: A review of the evidence. Menopause: Journal of the North American Menopause Society, 7(3), 184-192.

Ormel, J., & Wohlfarth, T. (1991). How neuroticism, long-term difficulties, and life situation change influence psychological distress: A longitudinal model. Journal of Personality and Social Psychology, 60(5), 744-755.

Ormel, J., Stewart, R., & Sanderman, R. (1989). Personality as a modifier of the life change-distress relationship. A longitudinal modeling approach. Social Psychiatry and Psychiatric Epidemiology, 24(4), 187-195.

Pajer, K. (1995). New strategies in the treatment of depression in women. Journal of Clinical Psychiatry, 56 (Suppl 2), 30-7.

Pardo, J. V., Pardo, P. J., & Raichle, M. E. (1993). Neural correlates of self-induced dysphoria. American Journal of Psychiatry, 150(5), 713-719.

Parry, B. L. (1989). Reproductive factors affecting the course of affective illness in women. Psychiatric Clinics of North America, 12(1), 207-220.

Pearlin, L. I. (1975). Sex roles and depression. In N. Datan, & L. H. Ginsberg (Eds), Life-Span Developmental Psychology: Normative Life Crises (pp. 191-207). New York: Academic Press.

Peselow, E. D., Dunner, D. L., Fieve, R. R., & DiFiglia, C. (1991). The prophylactic efficacy of tricyclic antidepressants--a five year followup. Prog Neuropsychopharmacol Biol Psychiatry, 15(1), 71-82.

Post, R. M. (1992). Transduction of psychosocial stress into the neurobiology of recurrent affective disorder. American Journal of Psychiatry, 149(8), 999-1010.

Poznanski, E. (1982). The clinical phenomenology of childhood depression. American Journal of Orthopsychiatry, 52(2), 308-313.

. (1976). J. A. Pritchard, & P. C. MacDonald Williams Obstetrics (15th ed.). New York: Appleton-Century-Crofts.

Punyahotra, S., Dennerstein, L., & Lehert, P. (1997). Menopausal experiences of Thai women. Part 1: symptoms and their correlates. Maturitas, 26(1), 1-7.

Radloff, L. S. (1977). The CES-D Scale: a self-report depression scale for research in the general population. Applied Psychological Measurement, 1(3), 385-401.

Roberts, R. E., Lewinsohn, P. M., & Seeley, J. R. (1991). Screening for adolescent depression: a comparison of depression scales. Journal of the American Academy of Child and Adolescent Psychiatry, 30(1), 58-66.

Roberts, R. E., Rhoades, H. M., & Vernon, S. W. (1990). Using the CES-D scale to screen for depression and anxiety: effects of language and ethnic status. Psychiatry Research, 31, 69-83.

Roberts, R. E., & Vernon, S. W. (1983). The Center for Epidemiologic Studies depression scale: its use in a community sample. American Journal of Psychiatry, 140(1 January), 41-46.

Roberts, R. E., Vernon, S. W., & Rhoades, H. M. (1989). Effects of language and ethnic status on reliability and validity of the Center for Epidemiologic Studies-Depression Scale with psychiatric patients. Journal of Nervous and Mental Disease, 177(10), 581-592.

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(4), 381-389.

Rohde, P., Lewinsohn, P. M., & Seeley, J. R. (1994). Are adolescents changed by an episode of major depression? Journal of the American Academy of Child and Adolescent Psychiatry, 33(9), 1289-98.

Rohde, P., Lewinsohn, P. M., & Seeley, J. R. (1990). Are people changed by the experience of having an episode of depression? A further test of the scar hypothesis. Journal of Abnormal Psychology, 99(3), 264-71.

Rohde, P., Lewinsohn, P. M., Tilson, M., & Seeley, J. R. (1990). Dimensionality of coping and its relation to depression. Journal of Personality and Social Psychology, 58(3), 499-511.

Rosenbaum, M., Lewinsohn, P. M., & Gotlib, I. H. (1996). Distinguishing between state-dependent and non-state-dependent depression-related psychosocial variables. British Journal of Clinical Psychology, 35(Pt 3), 341-58.

Rosner, B. (1986). Fundamentals of biostatistics. (2nd ed.). Boston: Duxbury Press.

Ross, C. E., & Mirowsky, J. (1984). Components of depressed mood in married men and women. American Journal of Epidemiology, 119(6), 997-1004.

Rothman, K. J. (1986). Modern epidemiology. Boston: Little, Brown.

Roy, A., Linnoila, M., Jimerson, D. C., Gold, P. W., & et al. (1986). Pre-menopausal and post-menopausal depressed women. Australian & New Zealand Journal of Psychiatry, 20(4), 464-469.

Rubinow, D. R., Hoban, M. C., Grover, G. N., Galloway, D. S., Roy-Byrne, P., Andersen, R., & Merriam, G. R. (1988). Changes in plasma hormones across the menstrual cycle in patients with menstrually related mood disorder and in control subjects. American Journal of Obstetrics and Gynecology, 158 (1 January), 5-11.

Rubinow, D. R., Schmidt, P. J., & Roca, C. A. (1998b). Estrogen-serotonin interactions: implications for affective regulation. Biological Psychiatry, 44(9), 839-50.

Rubinow, D. R., Schmidt, P. J., & Roca, C. A. (1998a). Hormone measures in reproductive endocrine-related mood disorders: diagnostic issues. Psychopharmacology Bulletin, 34(3), 289-90.

Rueda Martinez de Santos, J. R. (1997). Medicalization of menopause and public health. Journal of Psychosomatic Obstetrics and Gynaecology, 18(2), 175-180.

Sachar, E. J., Altman, N., Gruen, P. H., Glassman, A., Halpern, F. S., & Sassin, J. (1975). Human growth hormone response to levodopa. Relation to menopause, depression, and plasma dopa concentration. Archives General Psychiatry, 32(4), 502-3.

Sagsoz, N., Oguzturk, O., Bayram, M., & Kamaci, M. (2001). Anxiety and depression before and after the menopause (abstract only available). Archives Gynecol Obstet, 264(4), 199-202.

Saletu, B., Brandstätter, N., Metka, M., Stamenkovic, M., Anderer, P., Semlitsch, H. V., Heytmanek, G., Huber, J., Grünberger, J., Linzmayer, L., Kurz, Ch., Decker, K., Binder, G., Knogler, W., & Koll, B. (1995). Double-blind, placebo-controlled, hormonal, syndromal and EEG mapping studies with transdermal oestradiol therapy in menopausal depression. Psychopharmacology (Berl), 122(4), 321-9.

Saletu, B., Brandstätter, N., Metka, M., Stamenkovic, M., Anderer, P., Semlitsch, H. V., Heytmanek, G., Huber, J., Grünberger, J., Linzmayer, L., Kurz, Ch., Decker, K., Binder, G., Knogler, W., & Koll, B. (1996). Hormonal, syndromal and EEG mapping studies in menopausal syndrome patients with and without depression as compared with controls. Maturitas, *23*(1), 91-105.

Samuelsson, S. (1982). Life events and mental disorder in an urban female population. Acta Psychiatrica Scandinavica, Vol 65(Supplementum 299).

Schafran, L. H. (1996). Topics for our times: rape is a major public health issue. American Journal of Public Health, *86*(1), 15-16.

Schmidt, P. J., Nieman, L., Danaceau, M. A., Tobin, M. B., Roca, C. A., Murphy, J. H., & Rubinow, D. R. (2000). Estrogen replacement in perimenopause-related depression: a preliminary report. American Journal of Obstetrics and Gynecology, *183*(2), 414-420.

Schmidt, P. J., & Rubinow, D. R. (1991). Menopause-related affective disorders: A justification for further study. American Journal of Psychiatry, *148*(7), 844-52.

Scholle, S. H., Rost, K. M., & Golding, J. M. (1998). Physical abuse among depressed women. Journal of General Internal Medicine, *13*(Sept ), 607-613.

Schreiber, R. (1996). Understanding and helping depressed women. Archives of Psychiatric Nursing, *X*(3), 165-175.

Shea, M. T., Elkin, I., Imber, S. D., Sotsky, S. M., Watkins, J. T., Collins, J. F., Pilkonis, P. A., Beckham, E., Glass, D. R., Dolan, R. T., & Parloff, M. B. (1992). Course of depressive symptoms over follow-up: Findings from The National Institute of Mental Health Treatment of Depression Collaborative Research program. Archives General Psychiatry, *49*(10), 278-787.

Shea, M. T., & Hirschfeld, R. M. A. (1996). Chronic mood disorder and depressive personality. Psychiatric Clinics of North America, *19*(1), 103-120.

Sheehan, T. J., Fifield, J., Reisine, S., & Tennen, H. (1995). The measurement structure of the Center for Epidemiologic Studies Depression scale. Journal of Personality Assessment, *64*(3), 507-21.

Silberg, J. L., Heath, A. C., Kessler, R., Neale, M. C., Meyer, J. M., Eaves, L. J., & Kendler, K. S. (1990). Genetic and environmental effects on self-reported depressive symptoms in a general population twin sample. Journal of Psychiatric Research, *24*(3), 197-212.

Simon, G. E. (2000). Long-term prognosis of depression in primary care. Bull World Health Organ, 78(4), 439-45.

Simon, G. E., Katon, W., Rutter, C., Von Korff, M., Lin, E., Robinson, P., Bush, T., Walker, E. A., Ludman, E., & Russo, J. (1998). Impact of improved depression treatment in primary care on daily functioning and disability. Psychological Medicine, 28(3), 693-701.

Skodol, A. E. (1998). Personality and coping as stress-attenuating or -amplifying factors. In Adversity, stress, and psychopathology (pp. 377-389). New York, Oxford: Oxford University Press.

Smith, R. N., & Studd, J. W. W. (1994). Estrogens and depression in women. In R. A. Lobo (Ed), Treatment of the Postmenopausal Woman: Basic and Clinical Aspects, Chap. 11, (pp. 129-136). New York: Raven Press.

Somervell, P. D., Beals, J., Kinzie, J. D., Boehnlein, J., Leung, P., & Manson, S. M. (1993). Criterion validity of the Center for Epidemiologic Studies Depression Scale in a population sample from an American Indian village. Psychiatry Research, 47, 255-266.

Sowers, M. F., Crawford, S. L., Sternfeld, B., Morganstein, D., Gold, E. B., Greendale, G. A., Evans, D., Neer, R., Matthews, K., Sherman, S., Lo, A., Weiss, G., & Kelsey Jennifer. (2000). SWAN: a multicenter, multiethnic, community-based cohort study of women and the menopausal transition. In R. A. Lobo, J. Kelsey, & R. Marcus (Eds), Menopause: biology and pathobiology. San Diego, CA: Academic Press.

Stewart, D. E., & Robinson, G. E. (1995). Violence against women. In J. M. Oldham, & M. Riba (Eds), Review of Psychiatry (pp. 261-283). Washington, D. C.: American Psychiatric Press Inc.

Stommel, M., Given, B. A., Given, C. W., Kalaian, H. A., Schulz, R., & McCorkle, R. (1993). Gender bias in the measurement properties of the Center for Epidemiologic Studies Depression Scale (CES-D). Psychiatry Research, 49(3), 239-50.

Stoppard, J. M. (2000). Understanding depression: feminist social constructionist approaches. London & New York: Routledge.

Strickler, R. C., Borth, R., Cecutti, A., Cookson, B. A., Harper, J. A., Potvin, R., Riffel, P., Sorbara, V. J., & Woolever, C. A. (1977). The role of oestrogen replacement in the climacteric syndrome. Psychol Medicine, 7(4), 631-639.

Sturm, R., & Wells, K. B. (1995). How can care for depression become more cost-effective? JAMA, 273(1), 51-58.

Thomson, J. (1976). Double blind study on the effect of estrogen on sleep, anxiety and depression in perimenopausal women: preliminary results. Proc R Soc Med, 69(11), 829-30.

Tiggemann, M., Winefield, A. H., Winefield, H. R., & Goldney, R. D. (1991). The stability of attributional style and its relation to psychological distress. British Journal of Clinical Psychology, 30 (Pt 3), 247-55.

Tinkle, M. B., & Beaton, J. L. (1986). Toward a new view of science: implications for nursing research. in L. H. Nicoll (Ed), Perspectives on nursing theory (pp. 585-593). Boston, Toronto: Little Brown.

Treloar, A. E. (1974). Menarche, menopause, and intervening fecundability. Human Biology, 46(1), 89-107.

Treloar, A. E., Boynton, R. E., Behn, B. G., & Brown, B. W. (1967). Variation of the human menstrual cycle through reproductive life. International Journal of Fertility, 12(1 pt 2), 77-126.

Turner, R. J., & Noh, S. (1988). Physical disability and depression: a longitudinal analysis. Journal of Health & Social Behavior, 29, 23-37.

Ussher, J. M. (1992). Research and theory related to female reproduction: Implications for clinical psychology. British Journal of Clinical Psychology, 31(2), 129-151.

Utian, W. H. (1991). Menopause--a proposed new functional definition. Maturitas, 14(1), 1-2.

Utian, W. H. (1994). Ovarian function, therapy-oriented definition of menopause and climacteric. Experimental Gerontology, 29(3/4), 245-251.

Vollman, R. F. (1956). The degree of variability of the length of the menstrual cycle in correlation with age of woman. Gynaecologia, 142(5), 310-314.

Vollman, R. F. (1977). The menstrual cycle. Philadelphia: W. B. Saunders Co.

Walker, E. A., Gelfand, A., Katon, W. J., Koss, M. P., Von Korff, M., Bernstein, D., & Russo, J. (1999). Adult health status of women with histories of childhood abuse and neglect. American Journal of Med, 107(4), 332-9.

Walker, E., & Gelfand, A. (1997). Adult sequelae of childhood sexual victimization: New research findings. 20th Annual Pacific Northwest Conference for Primary Care Practitioners (Session D9). Seattle, WA: University of Washington Continuing Nursing Education.

Walker, E. A., & Katon, W. J. (1996). Editorial--Researching the health effects of victimization: the next generation. Psychosomatic Medicine, *58*, 16-17.

Walker, E. A., Katon, W. J., Hansom, J., Harrop-Griffiths, J., Holm, L., Jones, M. L., Hickok, L., & Jemelka, R. P. (1992). Medical and psychiatric symptoms in women with childhood sexual abuse. Psychosomatic Medicine, *54*, 658-664.

Walker, E. A., Torkelson, N., Katon, W. J., & Koss, M. P. (1993). The prevalence rate of sexual trauma in a primary care clinic. Journal of the American Board of Family Practice, *6*(5), 465-71.

Warner, V., Weissman, M. M., Fendrich, M., Wickramaratne, P., & Moreau, D. (1992). The course of major depression in the offspring of depressed parents: Incidence, recurrence, and recovery. Archives General Psychiatry, *49*(10), 795-801.

Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. Psychological Bulletin, *96*(3), 465-490.

Weissman, M. M. (1979). The myth of involuntional melancholia. JAMA, *242*(8), 742-4.

Weissman, M. M., Bland, R. C., Canino, G. J., Faravelli, C., Greenwald, S., Hwu, H.-G., Joyce, P. R., Karam, E. G., Lee, C.-K., Lellouch, J., Lépine, J.-P., Newman, S. C., Rubio-Stipec, M., Wells, J. E., Wickramaratne, P. J., Wittchen, H.-U., & Yeh, E.-K. (1996). Cross-national epidemiology of major depression and bipolar disorder. JAMA, *276*(4), 293-9.

Weissman, M. M., Sholomskas, D., Pottenger, M., Prusoff, B. A., & Locke, B. Z. (1977). Assessing depressive symptoms in five psychiatric populations: a validation study. American Journal of Epidemiology, *106*(3), 203-213.

Wells, K. B., Burnam, M. A., Rogers, W., Hays, R., & Camp, P. (1992). The course of depression in adult outpatients: Results from the Medical Outcomes study. Archives of General Psychiatry, *49*, 788-794.

Westhoff, C., Truman, C., Kalmuss, D., Cushman, L., Davidson, A., Ruline, M., & Heartwell, S. (1998). Depressive symptoms and Depo-Provera®. Contraception, *57*, 237-240.

Westhoff, C., Truman, C., Kalmuss, D., Cushman, L., Ruline, M., Heartwell, S., & Davidson, A. (1998). Depressive symptoms and Norplant® contraceptive implants. Contraception, *57*, 241-245.

Widom, C. S. (1999). Childhood victimization and the development of personality disorders. Archives of General Psychiatry, *56*(7), 607-8.

Widom, C. S. (1998). Childhood victimization: early adversity and subsequent psychopathology. In Adversity, stress, and psychopathology (pp. 81-95). New York, Oxford: Oxford University Press.

Wilbur, J., & Dan, A. J. (1989). The impact of work patterns on psychological well-being of midlife nurses. Western Journal of Nursing Research, 11(6), 703-716.

Williams, J. B. W., Spitzer, R. L., Linzer, M., Kroenke, K., Hahn, S. R., DeGruy, F. V., & Lavev, A. (1995). Gender differences in depression in primary care. American Journal of Obstetrics and Gynecology, 173(2), 654-9.

Winokur, G. (1997). All roads lead to depression: clinically homogeneous, etiologically heterogeneous. Journal of Affective Disorders, 45(1-2), 97-108.

Winokur, G. (1973). Depression in the menopause. American Journal of Psychiatry, 130(1), 92-93.

Wismont, J. M. (2000). The lived pregnancy experience of women in prison. Journal of Midwifery & Women's Health, 45(4), 292-300.

Wolk, S. I., & Weissman, M. M. (1995). Women and depression: an update. In J. M. Oldham, & M. Riba (Eds), Review of Psychiatry (pp. 227-259). Washington, D. C.: American Psychiatric Press Inc.

Woods, N. F. (1985). Employment, family roles, and mental ill health in young married women. Nursing Research, 34(1), 4-10.

Woods, N. F., & Mitchell, E. S. (1997). Pathways to depressed mood for midlife women: Observations from the Seattle Midlife Women's Health Study. Research in Nursing & Health, 20, 119-129.

Woods, N. F., & Mitchell, E. S. (1996). Patterns of depressed mood in midlife women: Observations from the Seattle Midlife Women's Health Study. Research in Nursing & Health, 19, 111-123.

Woods, N. F., Mitchell, E. S., & Mariella, A. M. (2000). Stress, neuroendocrine profiles and symptoms during the transition to menopause. Menopause: Journal of the North American Menopause Society, 7(6), 431.

Worsley, A., Walters, W. A. W., & Wood, E. C. (1977). Screening for psychological disturbance amongst gynaecology patients. Australian and New Zealand Journal of Obstetrics and Gynaecology, 17, 214-219.

Yin, R. K. (1989). Case study research: Design and methods. (Rev. ed.). Newbury Park, London: Sage.

Young, M. A., Fogg, L. F., Scheftner, W., Fawcett, J. A. H., & Maser, J. (1996). Stable trait components of hopelessness: baseline and sensitivity to depression. Journal of Abnormal Psychology, 105(2), 155-165.

Zich, J. M., Attkisson, C. C., & Greenfield, T. K. (1990). Screening for depression in primary care clinics: the CES-D and the BDI. International Journal of Psychiatry in Medicine, 20(3), 259-277.

Zonderman, A. B., Herbst, J. H., Schmidt Jr., C., Costa Jr., P. T., & McCrae, R. R. (1993). Depressive symptoms as a nonspecific, graded risk for psychiatric diagnoses. Journal of Abnormal Psychology, 102(4), 544-552.

**APPENDIX A****QUESTIONNAIRES and DATA COLLECTION INSTRUMENTS****Center For Epidemiologic Studies-Depression Scale (CES-D)****Health Report questionnaire items****Diary Medications****Menstrual Calendar Card**

### CENTER FOR EPIDEMIOLOGIC STUDIES-DEPRESSION SCALE

Below is a list of the ways you might have felt or behaved. Please circle the number that describes how often you have felt this way during the past week for each item.

0 = Rarely or none of the time (less than 1 day)

1 = Some or little of the time (1-2 days)

2 = Moderate amount of the time (3-4 days)

3 = Most or all of the time (5-7 days)

<u>During the past week:</u>	<u>&lt; 1 Day</u>	<u>1-2 Days</u>	<u>3-4 Days</u>	<u>5-7 Days</u>
1. I was bothered by things that usually don't bother me.....	0.....	1.....	2.....	3
2. I did not feel like eating; my appetite was poor.....	0.....	1.....	2.....	3
3. I felt that I could not shake off the blues even with help from my family or friends.....	0.....	1.....	2.....	3
4. I felt that I was just as good as other people.....	0.....	1.....	2.....	3
5. I had trouble keeping my mind on what I was doing.....	0.....	1.....	2.....	3
6. I felt depressed.....	0.....	1.....	2.....	3
7. I felt that everything I did was an effort.....	0.....	1.....	2.....	3
8. I felt hopeful about the future.....	0.....	1.....	2.....	3
9. I thought my life had been a failure.....	0.....	1.....	2.....	3
10. I felt fearful.....	0.....	1.....	2.....	3
11. My sleep was restless.....	0.....	1.....	2.....	3
12. I was happy.....	0.....	1.....	2.....	3
13. I talked less than usual.....	0.....	1.....	2.....	3
14. I felt lonely.....	0.....	1.....	2.....	3
15. People were unfriendly.....	0.....	1.....	2.....	3
16. I enjoyed life.....	0.....	1.....	2.....	3
17. I had crying spells.....	0.....	1.....	2.....	3
18. I felt sad.....	0.....	1.....	2.....	3
19. I felt that people dislike me.....	0.....	1.....	2.....	3
20. I could not "get going".....	0.....	1.....	2.....	3

## ITEMS FROM HEALTH REPORT QUESTIONNAIRES

**The next questions are about your use of medications:**

102. It is important for us to know what vitamins, minerals, medications and herbal or naturopathic preparations you take at least once a month. Please list all prescriptions and over the counter preparations you take on a regular basis (at least once a month). Think about what you took last month.

Write in None if you take no prescription or other preparations at least once a month.

Give the name, strength, amount, and how often you take a medicine or preparation and approximate date you started taking it.

	<u>Name</u>	<u>Strength</u>	<u>Amount &amp; Frequency</u>	<u>Date First Started (month/year)</u>
Example:	Premarin	.625 mg	one daily	10/96
Example:	Provera	5 mg	one days 1-12	10/96
Example:	Glyxase	3 mg	one twice a day	4/97
Example:	Advil	200 mg	two as needed	about 1992

	<u>Name</u>	<u>Strength</u>	<u>Amount &amp; Frequency</u>	<u>Date Started (month/year)</u>
a)	_____	_____	_____	_____
b)	_____	_____	_____	_____
c)	_____	_____	_____	_____
d)	_____	_____	_____	_____
e)	_____	_____	_____	_____
f)	_____	_____	_____	_____
g)	_____	_____	_____	_____
h)	_____	_____	_____	_____
i)	_____	_____	_____	_____
j)	_____	_____	_____	_____
k)	_____	_____	_____	_____
l)	_____	_____	_____	_____
m)	_____	_____	_____	_____

## ITEMS FROM HEALTH REPORT QUESTIONNAIRES

### Health Report questionnaire 1995

It is of increasing concern to health professionals what the effects of sexual harrassment, sexual abuse, and/or physical abuse and stress have on the health of women.

42a. **During the past year** have you experienced, in your opinion, any **sexual harrassment** either on the job or elsewhere?

No.....0

Yes.....1

42b. At **any time in your past**, have you ever experienced any:

	Yes	No
sexual abuse.....	1.....	0
physical abuse.....	1.....	0

A. If **Yes** to **either** item above, did this occur:

	Yes	No
during childhood (up to age 12).....	1.....	0
during adolescence (ages 12-20).....	1.....	0
during adult years (21 and over).....	1.....	0

### Health Report questionnaire 1997

38. **During the past year** have you experienced, in your opinion, any **sexual harrassment** either on the job or elsewhere?

yes.....1

no.....0

39. **During the past year**, have you experienced any:

	<u>Yes</u> .....	<u>No</u>
Sexual abuse.....	1.....	0
Physical abuse.....	1.....	0

**Health Report questionnaire 1999**

73. Have you ever been sexually assaulted, abused or molested?

Yes-----1

No-----0

74. **During the past year** did you experience, by your definition, any **sexual harassment** either on the job or elsewhere?

Yes---1

No----0

75. **During the past year**, did you experience any sexual abuse?

Yes-----1

No-----0

76. **During the past year**, did you experience any physical abuse?

Yes-----1

No-----0

**Seattle Midlife Women's Health Study**  
**Descriptive Statistics of Questionnaire Items Related to Abuse, Assault, Harassment**

**HEALTH REPORT QUESTIONNAIRE 1995**

**Item 42a. HARRAS95 Sexual Harassment Past Year**

		Frequency	Percent	Valid Percent	Cumulative Percent
No	0	128	90.1	90.8	90.8
Yes	1	13	9.2	9.2	100.0
	Total	141	99.3	100.0	
Missing	System	1	.7		
Total		142	100.0		

**Item 42b. SABUSE95 Sexual Abuse Any Past**

		Frequency	Percent	Valid Percent	Cumulative Percent
No	0	99	69.7	70.7	70.7
Yes	1	41	28.9	29.3	100.0
	Total	140	98.6	100.0	
Missing	System	2	1.4		
Total		142	100.0		

Eighteen (18) of these women reported the sexual abuse was as a child (13% of 142), and 23 were abused as teens (16% of 142).

**Item 42b. PABUSE95 Physical Abuse Any Past**

		Frequency	Percent	Valid Percent	Cumulative Percent
No	0	101	71.1	72.7	72.7
Yes	1	38	26.8	27.3	100.0
	Total	139	97.9	100.0	
Missing	System	3	2.1		
Total		142	100.0		

Thirteen (13) of these women reported the physical abuse was as a child (9% of 142), and 20 were abused as teens (14% of 142).

**Item 42b. A. CABUSE95 When Abuse-Childhood**

		Frequency	Percent	Valid Percent	Cumulative Percent
No	0	39	27.5	60.0	60.0
Yes	1	26	18.3	40.0	100.0
	Total	65	45.8	100.0	
Missing	8	75	52.8		
	System	2	1.4		
	Total	77	54.2		
Total		142	100.0		

**Item 42b. A. TABUSE95 When Abuse-Teen(Adolescence)**

		Frequency	Percent	Valid Percent	Cumulative Percent
No	0	32	22.5	49.2	49.2
Yes	1	33	23.2	50.8	100.0
	Total	65	45.8	100.0	
Missing	8	75	52.8		
	System	2	1.4		
	Total	77	54.2		
Total		142	100.0		

**Item 42b. A. AABUSE95 When Abuse-Adult**

		Frequency	Percent	Valid Percent	Cumulative Percent
No	0	30	21.1	45.5	45.5
Yes	1	36	25.4	54.5	100.0
	Total	66	46.5	100.0	
Missing	8	74	52.1		
	System	2	1.4		
	Total	76	53.5		
Total		142	100.0		

**HEALTH REPORT QUESTIONNAIRE 1997****Item 38. SEXHAR97 Any Sex Harass Job Or Elsewhere Pst Yr**

		Frequency	Percent	Valid Percent	Cumulative Percent
No	0	205	92.8	92.8	92.8
Yes	1	16	7.2	7.2	100.0
	Total	221	100.0	100.0	

**Item 39. SEXABS97 Any Sex Abuse Past Year**

		Frequency	Percent	Valid Percent	Cumulative Percent
No	0	219	99.1	99.5	99.5
Yes	1	1	.5	.5	100.0
	Total	220	99.5	100.0	
Missing	System	1	.5		
Total		221	100.0		

**Item 39. PHYABS97 Any Physical Abuse Past Year**

		Frequency	Percent	Valid Percent	Cumulative Percent
No	0	217	98.2	98.6	98.6
Yes	1	3	1.4	1.4	100.0
	Total	220	99.5	100.0	
Missing	System	1	.5		
Total		221	100.0		

**HEALTH REPORT QUESTIONNAIRE 1999****Item 73. ABUSED99 Ever Sexually Assaulted, Abused Or Molested**

		Frequency	Percent	Valid Percent	Cumulative Percent
No	0	130	61.3	67.4	67.4
Yes	1	63	29.7	32.6	100.0
	Total	193	91.0	100.0	
Missing	System	19	9.0		
Total		212	100.0		

**Item 74. SEXHAR99 Experienced sexual harassment in past year**

		Frequency	Percent	Valid Percent	Cumulative Percent
No	0	186	87.7	95.9	95.9
Yes	1	8	3.8	4.1	100.0
	Total	194	91.5	100.0	
Missing	System	18	8.5		
Total		212	100.0		

**Item 75. SEXABS99 Experienced sexual abuse in past year**

		Frequency	Percent	Valid Percent	Cumulative Percent
No	0	192	90.6	99.0	99.0
Yes	1	2	.9	1.0	100.0
	Total	194	91.5	100.0	
Missing	System	18	8.5		
Total		212	100.0		

**Item 76. PHYABS99 Experienced physical abuse in past year**

		Frequency	Percent	Valid Percent	Cumulative Percent
No	0	191	90.1	98.5	98.5
Yes	1	3	1.4	1.5	100.0
	Total	194	91.5	100.0	
Missing	System	18	8.5		
Total		212	100.0		



# MENSTRUAL CALENDAR CARD

Year \_\_\_\_\_

MENSTRUAL CYCLE RECORD

IDNO \_\_\_\_\_

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
JAN																																
FEB																																
MAR																																
APR																																
MAY																																
JUN																																
JUL																																
AUG																																
SEP																																
OCT																																
NOV																																
DEC																																

For every day you spot or bleed enter an S or B in the appropriate square. Record a 1, 2, 3 or 4 next to every B day.  
 (1: light flow, 2: moderate, 3: heavy, 4: very heavy/bleeding)  
 For any month no bleeding or spotting occurs write in NO BLEEDING.  
 If you forget to record for a month write in FORGOT TO RECORD.

RECORD any unusual events including surgery (cesarean), illnesses and medications, started or stopped, include major events regarding health, work, residence, relationships, financial, legal matters, parenting. If you start, stop or change hormones please record this with the date.

DATE EVENT/MED. CHANGE

---

---

---

---

---

---

---

---

Address or phone change:

Return To:

Center for Women's Health Research  
 School of Nursing (206) 543-1414  
 University of Washington Box 357261  
 Seattle, WA 98195-7261

**APPENDIX B  
CONSENT FORMS**

UNIVERSITY OF WASHINGTON  
SCHOOL OF NURSING  
Center for Women's Health Research  
Midlife Women's Health and Health-Seeking Behavior  
Consent Form

Principal Investigator: Nancy F. Woods, R.N., Ph.D.  
Professor, Department of Parent and  
Child Nursing  
Phone: 543-8221

Co-investigator: Ellen S. Mitchell, A.R.N.P., Ph.D.  
Research Assistant Professor, Department  
of Parent and Child Nursing  
Phone: 543-1414

Co-investigator: Martha J. Lentz, R.N., Ph.D.  
Research Assistant Professor, Department  
of Physiological Nursing  
Phone: 543-1414

Purpose and Benefits

The purpose of this study is to find out about the kind of symptoms women in midlife experience over a three year time period and whether these symptoms are related to women's health practices, their past medical history, present health, stress in their lives, and their attitudes about women, aging, and their roles in society. In addition, we hope to find out how women cope with stress, how they express their anger, and their feelings about wellness and health. The benefits of this study will be a better understanding for women in general and for health professionals of how common these symptoms are and what factors influence women's health in midlife.

Procedures

You will be asked to participate in a two-hour interview in a place of your choice about your past medical history, current health practices, present health, your attitudes about women and their roles in society, your thoughts and feelings about sex, and about stress in your life. You will be asked questions like, "Have you ever used birth control pills?", "Have you had a death in your family in the past six months?", and "Have you had a change in sexual thoughts in the past year?" You will be asked to keep track of your health in a daily diary for about three months, and to keep a menstrual calendar for a total of three years. The diary and calendar will require 5-10 minutes at the end of each day. The diary asks about symptoms you may experience, how stressful or rewarding was your day, how you ..

coped with a stressor, and about your health practices such as drinking alcohol, smoking, or having sexual intercourse that day. The menstrual calendar requires that you circle the date when you had bleeding or spotting. In approximately two weeks after beginning the diary and calendar the interviewer will contact you by phone to see if you have any questions.

At the end of one year and at two years you will again be asked to keep a similar diary and menstrual calendar. At the end of one year, two years, and three years you will be asked to complete a health update questionnaire and a 30 minute telephone interview. The calendar, diary, and health update will be mailed to you. The second and third diary will be for about three months each and be the same as the first diary. In 7-14 days after everything is mailed to you an interviewer will call you for the telephone interview which will take about thirty minutes and will include questions about your roles and responsibilities, how you describe yourself, and about your self care. The health update will ask about changes in your social and economic resources, life events, menstrual cycle, health status, and health practices during the past year.

In summary, you will complete one in-person interview, three health diaries, three health updates and three phone interviews each a year apart, and a menstrual calendar for three years. Each time you return a completed diary you will be paid \$25 for a total of \$75.

In addition, during this study, with your permission, you may be contacted by other investigators or their staff associated with the Center for Women's Health Research to ask if you are interested in participating in other Center studies.

#### Risks, Stress, or Discomfort

Some of the questions asked might make you feel anxious or embarrassed. During all parts of the study, you are free not to participate and to withdraw from the study at any time without penalty or loss of benefits to which you are otherwise entitled.

Any information you provide will be kept in confidence. Your questionnaire, diary, calendar, and telephone interview data will only be identified by number and kept in a locked file cabinet. Your name will not be used in any reports and it will not be attached to your answers. Only the study personnel will have access to the data. All data will be kept indefinitely.

---

Coinvestigator & Project Director

**Participant's Statement**

The study described above has been explained to me, and I voluntarily consent to participate in this activity. I have had an opportunity to ask questions and understand that future questions I may have about the research or about subjects's rights will be answered by the investigators listed above.

\_\_\_\_\_  
Signature of Participant / Date

\_\_\_\_\_  
Signature of Interviewer / Date

I give permission for my name, address, and phone number to be released to other investigators associated with the Center for Women's Health Research. I understand that I may be contacted to participate in other Center studies.

Yes \_\_\_\_\_

No \_\_\_\_\_

\_\_\_\_\_  
Signature of Participant / Date

Copies to:      Participant  
                    Investigator Files  
1/1/91

DO NOT PUT YOUR ID NO ON THIS PAGE

We know our study requested participation for 3 years only. However, many of you have not yet experienced any menopausal changes and for others you are in the transition time. We would like to continue following you until you have completed menopause since very little is known about what women experience as they approach and go through menopause.

Would you be willing to continue to work with us on a yearly basis and continue to provide information? Please indicate with a check mark ALL that you consent to:

- keeping a menstrual calendar that is mailed back yearly.
- filling out a questionnaire yearly similar to the one today.
- keeping a diary once a year that has fewer items than the ones already done in previous years.

PLEASE CHECK ONLY ONE OF THE OPTIONS BELOW:

- I would be willing to do those things checked above without pay.
- I would only be willing to do those things checked above if I were paid.
- I would NOT be willing to be contacted about any of the above.

I understand that I am free not to participate and to withdraw from any or all portions of this study at any time without penalty or loss of benefits to which I am otherwise entitled.

\_\_\_\_\_  
SIGNATURE

If you discontinue participation and wish to withdraw please notify the Center for Women's Health Research. Again, any information you provide will be kept in confidence. Your information will only be identified by number and kept in a locked file cabinet. Your name will not be used in any reports and it will not be attached to your answers. Only the study personnel will have access to the data. All data will be kept indefinitely.

Please remember to mail these questionnaires and your old calendar (if you will no longer continue recording) back in the enclosed prepaid envelope.

UNIVERSITY OF WASHINGTON  
SCHOOL OF NURSING  
Center for Women's Health Research  
Consent Form

Principal Investigator: Nancy F. Woods, R.N., Ph.D.  
Professor, Department of Parent Child  
Nursing  
Phone: 543-1414

Co-Investigator &  
Project Director Ellen S. Mitchell, A.R.N.P., Ph.D.  
Research Associate Professor  
Department of Parent Child Nursing  
Phone: (206) 543-1414

Purpose and Benefits

The purpose of this study is to extend the Midlife Women's Health Study according to the consent you gave at 36 months, that is, at the end of the original study. The benefits of this study will be increasing the information about longitudinal experiences (from year to year) of women in midlife approaching menopause. You are part of the only study in the United States that takes such a broad look at the health of women over time. We hope to continue yearly data collection for a number of years since very little is known about the health and lives of women in your age group.

Procedures

The calendar cards and Health Report Questionnaires are unchanged.

The diary is similar to the one you completed for the Midlife Women's Health Study, with the following changes:

1. Due to funding limitations, the "thank you" check is now only \$10.
2. This diary is shorter--only 6 weeks--and has fewer items to mark every day.
3. Instead of recording or circling numbers you will fill in a "bubble" or circle with pencil or black or blue ballpoint pen.
4. You will write in the master list of medications/vitamins and minerals on the back cover of the diary booklets yourself when you receive it in the mail. This list is not already written in for you.
5. Also due to funding limitations, there will be no planned phone contact.
6. Lastly because of funding limitations, everyone who is participating is receiving the mailings at the same time of year with the calendar. If you are completing both calendar and Health Report Questionnaires, you will return them both in the same envelope to economize on postage costs.

Please circle again YES or NO for those materials you would be willing to complete now and in the future.

\_\_\_\_\_ I am NOT willing to be contacted further.

YES NO Calendars

YES NO Health Reports (the questionnaires that update your health for the past year)

YES NO Diary (redesigned to be optically scannable and SHORTER)

Risks, Stress, or Discomfort

During this study you are free not to participate and to withdraw at any time without penalty or loss of benefits to which you are otherwise entitled.

Any information you provide will be kept in confidence. Your data will only be identified by number and kept in a locked file cabinet. Your name will not be used in any reports and it will not be attached to your answers. Only the study personnel will have access to the data. All data will be kept indefinitely.

\_\_\_\_\_  
Ellen S. Mitchell, A.R.N.P., Ph.D.  
Principal Investigator

Participant's Statement

The study described above has been explained to me, and I voluntarily consent to participate in this activity. I have had an opportunity to telephone the office and ask questions. I understand that future questions I may also have about the research or about subjects' rights will be answered by the investigators listed above.

\_\_\_\_\_  
Signature of Participant / Date

Please return one copy of this signed consent form with your completed materials. Thank you.

a:MHWEXTCN  
revised 10/31/94

**UNIVERSITY OF WASHINGTON  
CONSENT FORM**

**SCHOOL OF NURSING  
Center for Women's Health Research  
Menopausal Transition: Biobehavioral Models of Symptoms**

Principal Investigator: Nancy F. Woods, R.N., Ph.D., F.A.A.N.  
Professor, Dept. of Family & Child Nursing  
Director, Center for Women's Health Research  
Phone: 543-4090

Co-investigator & Project Director: Ellen S. Mitchell, A.R.N.P., Ph.D.  
Associate Professor  
Dept. of Family & Child Nursing  
Phone: 543-4090

**PURPOSE AND BENEFITS**

The purpose of this study is to learn how symptoms, hormones, and stress relate to each other in women going through the menopausal transition. Participation in the study is for four years. The benefits of this study will be a better understanding for women in general and for health professionals about factors that influence the transition to menopause.

**PROCEDURES**

You are asked to participate in a 2 hour interview and orientation session in a place of your choice. The study will be explained, instructions for keeping the diary and calendar will be reviewed, a list of your current medications and vitamins will be made, and the procedure for collecting and freezing urine specimens will be explained. Measurement of your height, weight, and waist and hip circumference will be obtained. In addition, you will be asked questions about your general health such as past history, family history and any current health problems. Finally, questions about sleep, eating, alcohol intake and smoking will be asked.

You are asked to keep the same 3 records that you have been doing in the previous study. These records are the yearly health update questionnaire, the menstrual calendar, and the diary. The menstrual calendar will be kept for four years. When you record in the diary will be different from before. Approximately once a month you are asked to keep the diary for 3 days. The days you keep the diary will be around the time you do your urine collection.

Once a month on arising you are asked to collect your urine in a plastic container coated with a special preservative and then to freeze it in your home freezer. Someone from the study will collect these specimens about every two months. Your menstrual calendar and health update questionnaire will be mailed at the end of each year.

Finally, once a year we will measure your blood pressure, pulse, weight, height and hip circumference.

**RISKS, STRESS, OR DISCOMFORT**

Some of the questions in the initial interview, diary, or health update might cause you some discomfort or stress. You will be asked questions like, "How would you describe your current level of sexual desire?", "During the past year have you experienced, in your opinion, any sexual harassment either on the job or elsewhere?", "How much do you agree with the statement: At times I think I am no good at all." "Did you ever have a period of 2 or more weeks when you were depressed?" and "Have you ever wanted to cut down on your drinking?" You are free not to answer any questions you do not wish to answer.

**OTHER INFORMATION**

You will receive \$200 each year for a total of up to \$800 for four years if all data are collected. You will be paid once a year. The payment will be divided as follows:

Monthly 3-day diary	= \$75/year
Monthly urine collection	= \$75/year
Health update yearly	= \$25/year
Calendar for one year	= \$25/year

All information will be kept confidential. All data will be coded with a number and kept in a secure area. Your name will not be used in any reports and will be kept separate from the data you provide. All data will be kept indefinitely.

During all parts of the study you are free not to participate and to withdraw at any time without penalty or loss of benefits to which you are otherwise entitled.

*Ellen S. Mitchell* 10/15/96  
 Ellen S. Mitchell / Date  
 Co-investigator & Project Director

**Participant's Statement**

The study described above has been explained to me, and I voluntarily consent to participate in this activity. I have had an opportunity to ask questions and understand that future questions I may have about the research or about subjects' rights will be answered by the investigators listed above.

\_\_\_\_\_  
 Signature of Participant / Date

Copies to: Participant  
 Investigator's file

nwh2con.doc  
 10/11/96/bal

**APPENDIX C**

**DSM-IV CRITERIA FOR MAJOR DEPRESSIVE EPISODE**

### Criteria for Major Depressive Episode

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

- (1) Depressed mood most of the day, nearly every day
- (2) Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day
- (3) Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day
- (4) Insomnia or hypersomnia nearly every day
- (5) Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)
- (6) Fatigue or loss of energy nearly every day
- (7) Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick)
- (8) Diminished ability to think or concentrate, or indecisiveness, nearly every day
- (9) Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

[Exclusions to the diagnosis have been omitted here.]

American Psychiatric Association Committee on Nomenclature and Statistics (1994). Mood disorders. In Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) Fourth ed., (p. 327). Washington, D.C.: American Psychiatric Association.

**Vita**

**Name:** Anne Marie Mariella

**Birth:** October 13, 1950  
Evanston, IL

**Education:** Loyola University  
Rome Center Campus, Rome, Italy  
Lake Shore Campus, Chicago, IL  
Lewis Towers Campus, Chicago, IL

Princeton University  
Princeton, NJ

University of Wisconsin  
Madison, WI

Columbia University  
New York, NY

Boston University  
Boston, MA

University of Illinois at the Health Sciences Center  
Chicago, IL

University of Washington  
Seattle, WA

**Degrees:** BA, Anthropology  
Princeton University, 1972

BSN  
Columbia University, 1975

MS, Nursing Science  
University of Illinois, 1982

PhD, Nursing Science  
University of Washington, 2001