Wittmann, W. W. (1988). Multivariate reliability theory: Principles of symmetry and successful validation strategies. In J. R. Nesselroade & R. B. Cattell (eds.), *Handbook of multivariate experimental psychology*. New York: Plenum.

Wohlwill, J. F. (1973). *The study of behavioral development*. New York: Academic Press.

Zevon, M. A. & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology, 43*, 111–122.

Zuckerman, M. (1983). The distinction between trait and state scales is not arbitrary: Comment on Allen and Potkay's 'On the arbitrary distinction between traits and states.' *Journal of Personality and Social Psychology, 44*, 1083–1086.

ACKNOWLEDGEMENTS

The research reported here and participation in the conference involved in the production of this volume was supported by the MacArthur Foundation Research Program on Successful Aging (MFRPSA). We deeply appreciate the stimulating interactions and discussions we have had with our MFRPSA colleagues concerning the issues discussed herein. We are also grateful to Julia Mullen, Terri Orbuch, and Joe Donato for their help in conducting the Cornwall Manor study.

4 Now you see it, now you don't – some considerations on multiple regression

GEORGE W. BROWN, TIRRIL O. HARRIS AND LOUISE LEMYRE

There is always a temptation to shrink from the difficult task of grounding identifying assumptions in empirical and theoretical knowledge and to select them instead on grounds of statistical convenience or simplicity. This is a temptation that is worth resisting if we wish to arrive at a real understanding of the phenomena we are studying. In this as in other decisions associated with our statistical methods, it is a snare and a delusion to seek out 'automatic' techniques that can be applied without careful consideration or that are neutral with respect to the substance of our theories. It is not enough to learn the etiquette book. The reason (or lack of reason) for its rules must be understood as well.

Herbert A. Simon (1979)

STATISTICS AND THE ANALYSIS OF SURVEY MATERIAL

It may be useful to begin with a small amount of history. This chapter derives from a series of seminars held in London to stimulate discussion of research issues in social psychiatry and was deliberately couched in somewhat provocative terms (hence its title). In a longer view it relates to a certain lingering skepticism about the usefulness of complex multivariate statistics – at least on the part of the first two of us. To be provocative: in our own field of social psychiatry we know of no important findings whose dissemination required such statistics; we suspect that most discoveries have not owed their origin to such statistics, and the statistics may at times actually have prevented important insights. Finally there is a danger that in an essentially multidisciplinary field they will keep many from comprehending potentially significant results or from viewing them sufficiently critically. Something of this general unease was conveyed by one of us in an earlier seminar in the present series in the context of discussing path analysis:

> In practice there has been a tendency to use it to test a limited range of hypotheses rather than to come to terms with the plethora of possibilities inherent in a complex data set. This practice is worrying and runs counter to the spirit of the Durkheim–Lazarsfeld tradition of analysis and often appears to

reflect a misunderstanding of the delicate balance between getting and testing ideas, between the imaginative and the critical in research (Medawar, 1969). There is an element of this interplay in path analysis in the sense that it might well reveal that certain paths can be excluded, but the possibilities for exploration are limited and I doubt whether its use alone will usually have the necessary depth of contact with material to provide an effective mode of analysis. Path coefficients obtained more or less routinely from a computer-based programme are no substitute for getting to know material by puzzling over it in its innumerable manifestations for months, if not years. It does not, for example, supply the all-important partial tables of the traditional Lazarsfeldian approach. Given the flawed measures, the emphasis on 'factors' and the primitive state of knowledge in most fields of inquiry, path analysis is probably best seen primarily as an aid to exploring material and as providing a useful method of summarizing the results of a broader and more complex analysis.

(Brown, 1989: 308)

In general we are impressed by the ease with which significant effects are easily missed even when working close to data in tabular or diagrammatic form, and how because of this it is highly desirable for a worker to give him or herself the opportunity of arriving at the same insight from as many different directions as possible. And this may only come at times from muddling through data, moving from one thing to another as ideas occur, and reworking analyses over and over again as new factors emerge.

However, we are equally impressed by the elegance and succinctness of a regression analysis done well, when the author knows his or her data thoroughly and has clearly looked at the material in many ways and explored in depth possible counter interpretations of the core results of the multivariate analysis. The perspective we espouse is then one of multi-methods: and the reader should be notified early on that our final somewhat critical conclusion relates far more to shortcomings in analysis than in the technique of multiple regression itself. In this spirit we decided to take a critical finding from our own research dealing with the aetiology of depression and ask what the likely consequences would have been if instead of following the traditional, basically tabular, methods of survey analysis we had employed multiple regression. We had, in fact, three questions:

1. whether we would have arrived at our original conclusions about the aetiology of depression (one that has informed a good deal of our research over the last 10 years) if we had used (as many have advised) multiple regression techniques;
2. whether, even if we had stuck to our traditional methods, we would have gained additional insights by using such techniques as well;
3. whether under such circumstances there was a chance of being led astray about what was going on.

A second issue emerged more prominently as the exercise progressed. We had held (perhaps with more uncertainty) equally unfashionable views about another topic: the advantages that accrued from seeing psychiatric disorder in categorical rather than continuous terms. The choice between the two perspectives cannot be reduced to questions about the fundamental nature of phenomena such as depression. We have been ready enough to recognize that the latter appear to be continuous – and, indeed, have argued that the distinction between 'psychotic' and 'neurotic' depression might best be seen in these terms. (We hasten to add that this is hardly an original view.) However, this is by no means the same as concluding that it is useful to *analyze* depression in such terms. Effective theory will not necessarily emerge from such verisimilitude: despite the apparent continuous nature of depression it may only be possible for certain insights to emerge when its apparent continuous nature is arbitrarily divided into categories. Arbitrary here does not mean that the categorization is unreliable, nor that some intuitive feel about the relevance of the divisions is missing. It merely means that there is no obvious justified 'break' in the continuity of the phenomenon. Justification therefore must come from its usefulness, theoretically and practically. Multiple regression, of course, encourages us to see depressive phenomena in continuous terms because of the potential increase in predictive power of continuous measures and the issue is an inevitable backdrop to the present exercise. The contrasting advantages of using categorical measures expressed as dummy variables in multiple regression are much less often recognized. There is, for example, no problem in using unstandardized regression coefficients in this way to obtain a direct sense of the relative importance of independent variables. This is because they are comparable in the sense of scales reflecting presence or absence (Achen, 1982: 70–1): we are simply asking whether the presence of X has a greater impact than the presence of W upon factor Y. This is non-controversial. The basic problem concerns the evaluation of the different results that can emerge in using a continuous measure as against a categorical measure as the dependent variable. Not only is there some dispute as to whether such a procedure is statistically viable (Hellevik, 1984; Schroeder, Sjoquist & Stephan, 1986), but, as will emerge, the picture presented seems to change even within the same data set.

THE BASIC RESULT IN CATEGORICAL TERMS

The aetiological model of depression to be explored was derived from research which approached the issue in the spirit of giving feed-back to clinicians about the psychosocial origins of that disorder. Like clinicians, it therefore adopted a categorical approach. The early research

Table 4.1. *Basic vulnerability result: vulnerability factor, presence of a provoking agent and onset of caseness depression (303 Islington women excluding cases of depression at first interview)*

Provoking agent	Vulnerability		
	Yes	No	
	Percentage onset case		
Yes	*29* (28/95)	*4* (2/55)	$p < 0.001$*
No	*2* (1/56)	*1* (1/97)	NS

* Significance level of at least 0.05.
NS, Not significant.

established that certain *provoking agents* determine when an episode of depression will take place. Most significant are severely threatening events, usually involving a loss or disappointment, although major ongoing difficulties can also play a role (Brown & Harris, 1978*a*; Brown, Bifulco & Harris, 1987). At a conservative estimate about 80% of episodes of depression in the general population are brought about by such an event or difficulty. However, the chance of such an event bringing about a depressive disorder appears to be considerably influenced by the presence of *vulnerability* factors, such as lack of support. In the original enquiry carried out in Camberwell in South London we concluded that lack of an intimate tie with a husband, having three or more children under 14 living at home, and loss of mother before 11 act as vulnerability factors for women – that is, that they increased risk of depression only in the presence of a provoking agent (Brown & Harris, 1978*a*; 181). (Lack of employment only related to increased risk in the presence of other vulnerability factors.)

Table 4.1 illustrates this effect in terms of a contingency table in which the proportions experiencing the dependent variable are expressed in terms of the four possible combinations of two dichotomous independent variables. The data are drawn from a recent longitudinal study in Islington where the index of vulnerability is based on measures of core relationships made at the time of first contact with the women involved – that is, before the occurrence of any severe events (which made up the great majority of the provoking agents) and before any onset of depression. The index representing vulnerability (negative elements in core relationships) was derived from an analysis of 70 measures dealing with negative interaction with husband, lack of primary quality in the relationship with husband, security-diminishing characteristics of

housewife role (a measure largely reflecting shortcomings in the practical and financial help provided by husband) and negative interaction with children (see Brown, Bifulco & Andrews, 1990 for details).

As already intimated, some understanding of the measure of depression used in the enquiry is important for our argument.

In the last 20 years, a number of standardized clinical-type interviews have been developed, using questions aimed to establish the presence of symptoms of a type and severity typically encountered in out-patient and hospital practice. Throughout our research we have used a shortened version of the Present State Examination (PSE) to collect material about basic symptomatology (Wing, Cooper & Sartorius, 1974). There is also in common use the *Schizophrenia and Affective Disorder Schedule* (SADS) (Endicott & Spitzer, 1978), and the *Clinical Interview Schedule* (Goldberg *et al.*, 1970). While each use somewhat different criteria for the inclusion of symptoms, they agree broadly, at a symptomatic level, on the range and severity of symptoms judged to be of psychiatric relevance.

The PSE was used, as designed, to collect symptoms for one month, and its use extended to cover the 12 months before interview in order not to miss the onset of cases earlier in the year, which are crucial for the study of the aetiological role of life events and difficulties (Brown & Harris, 1978*a*). The interviewer dated the onset as accurately as possible within the year, if necessary using events such as move of house as anchor points.

Interviewers, who were permanent members of the research team, were trained in the use and administration of the shortened version of the ninth edition of the Present State Examination (Wing *et al.*, 1974). Although initially designed to be administered by clinicians, there is now extensive evidence to support its use in social surveys by trained lay interviewers (see, e.g. Cooper *et al.*, 1977; Wing, Henderson & Winckle, 1977). Details of the clinical interview, as well as basic information on psychiatric conditions likely to be encountered in community surveys of this kind and on the identification and training and its use is given in Brown, Craig & Harris (1985).

The establishment of an acceptable caseness threshold is always the crux of plausibility in epidemiological psychiatric research using a categorical approach. The threshold used in our surveys has aimed to reflect current psychiatric practice; it was thus deliberately designed to contrast 'cases', who would have syndromes comparable to those of women seen in outpatient clinics, with 'borderline cases'. The latter have symptoms that are not sufficiently typical, frequent, or intense to be rated as cases, but still are more than odd isolated symptoms. There are also women with psychiatric symptoms such as fatigue, sleep disorder, and nervous tension which are not sufficient to warrant even a

borderline case rating. The original caseness threshold was established by two psychiatrists (John Cooper and John Copeland) when research began in 1969, but essential to the whole procedure has been the development of reference examples of cases and borderline cases for different diagnostic groups. In the general population, cases of depression and anxiety or phobic states are much the most frequent, but there are also categories for obsessional, tension, alcoholic, and drug-dependent conditions. This diagnostic system, which can now be applied using a simple computer algorithm in so far as it involves depressive and anxiety/phobic conditions, has been described extensively elsewhere, and has been shown to have good inter-rater reliability; there is also good evidence for its construct validity in the context of aetiological research (Brown & Harris, 1978*a*; Finlay-Jones *et al.*, 1980; Finlay-Jones and Brown, 1981; Brown & Prudo, 1981; Prudo, Brown, Harris & Dowland, 1981).

The following checklist has been shown statistically to underlie the clinical criteria for depression (Finlay-Jones *et al.*, 1980) and, as will be seen, has been emphasized in the present exercise.

For the diagnosis of a *case of depression*, both A and B must be present:

(A) depressed mood;

(B) four or more of the following symptoms: hopelessness, suicidal plans or attempts, weight loss, early waking, delayed sleep, poor concentration, neglect due to brooding, loss of interest, self-depreciation, and anergia.

In practice, many other symptoms covered by the PSE are also to be found present – on average about 19 in total.

The two criteria for the diagnosis of a *borderline case of depression* are:

(A) depressed mood;

(B) between one and three of the symptoms listed above.

Major syndromes are treated non-hierarchically. This permits an anxiety state to be rated separately from a depressive disorder and allows a subject to be characterized by separate diagnoses at different levels of severity – for example, case depression: borderline case anxiety. At this point we should perhaps state again that the divisions we have used are essentially arbitrary in the sense that the range of severity of the phenomonon appears to be continuous. Justification for the divisions must depend on what follows from making them.

Since we were concerned with the onset of a new caseness depressive condition Table 4.1 excludes all women suffering from such a condition at the time of first interview (50 women) and includes only the 303 women followed up at a second interview one year later. A total of 32

Table 4.2. *Vulnerability factor, presence of a provoking agent and onset of borderline caseness of depression (271 Islington women excluding cases of depression at first interview and onsets in follow-up years)*

Provoking agent	Vulnerability		
	Yes	No	
	Percentage onset borderline case		
Yes	*22* (15/67)	*15* (8/53)	NS
No	*2* (1/55)	*3* (3/96)	NS

Cases (shown in Table 4.1) are excluded from the denominators.

of the 303 women developed caseness of depression at some time in the follow-up year and, as can be seen, 30 of the 32 had experienced a provoking agent before onset (29 a severe event and 1 a major difficulty only).

We will concentrate on this basic result concerning onset of caseness. However, we will also refer to a second result that has now been replicated in several studies and begins to provide further evidence for the kind of justification we called for earlier for approaching depression in such categorical terms. Table 4.2 shows that the vulnerability effect does not occur for those developing a borderline case depressive condition: see the top row of Table 4.2 (women developing case conditions are excluded). Perhaps the most plausible interpretation is that women are not entirely protected from depression by the lack of vulnerability – but, given the presence of a provoking agent, they are likely to develop a milder, borderline case, rather than case condition, thus accounting for the high rate of the milder borderline conditions where there is a provoking agent but no vulnerability (17% in Table 4.2).

INTERACTIVE EFFECTS

There has been a good deal of controversy over the last 10 years about the analysis of the kind of interactive effect shown in Table 4.1, and some brief reference to it is perhaps in order. The controversy originated from two papers written by members of the MRC Social Psychiatry Research Unit about the basic results of the original Camberwell enquiry (Tennant & Bebbington, 1978; Bebbington, 1980). (See also our replies: Brown & Harris, 1978*b*; 1980.) No-one involved

in the debate has questioned the usefulness of dealing with depression in categorical terms. The key issue is whether lack of support (or any other 'vulnerability' factor) has an independent effect or only increases risk of depression in the presence of a provoking agent. Tennant and Bebbington concluded on the basis of our material that 'the social variables involved do not segregate into "vulnerability factors" and "provoking agents". Both categories of variables affect the risk of disorder independently of the other' (Tennant & Bebbington, 1978: 574); and that 'vulnerability factors have an effect somewhat similar to the separate provoking effect of life events' (p. 573). In other words they assert that vulnerability factors act statistically as provoking agents.

One problem with Tennant & Bebbington's criticism is their failure to appreciate that certain assumptions about the world are typically already built into statistical techniques; whether or not it is reasonable to go along with these assumptions cannot then be settled by statistical arguments depending on these very assumptions. Different statistical procedures handle interactive effects differently. It has been traditional in the social sciences to see the kind of result illustrated in Table 4.1 in terms of so-called additive interaction. Very briefly this requires first calculating the independent contributions of the two main variables. Thus in Table 4.1 the independent effect of the provoking agent is 4% (the effect of the provoking agent when vulnerability is not present) minus 1% (the expected rate of depression without either the provoking agent *or* vulnerability factor). Following the same logic the independent effect of the vulnerability factor is 2%–1%. It follows that the excess or interactive effect of provoking agent *and* vulnerability factor when occurring together is 24%: i.e., $29-((4-1)+((2-1)+1)$, the final 1% relating to the expected rate when neither factor is present. This is quite sufficient to establish the presence of a vulnerability effect. (It also follows exactly the procedure for establishing regression coefficients where there is a dichotomous dependent variable expressed, as for caseness, in terms of a proportion as in Table 4.1: Hellevik, 1984.)

However, there is an alternative way of considering interaction in which effects are multiplied rather than added – not to be confused with the multiplication of dummy independent variables in multiple regression (see Brown & Harris, 1986*a*: 151–2). The problem is that the two approaches can give quite different answers concerning the presence of interactive effects, and in the Camberwell material (but not in the present data) this multiplicative approach showed a complete absence of interaction and thereby lack of support for a distinctive vulnerability effect.

The original paper by Tennant & Bebbington failed to recognize that the ordinary log-linear approach they employed reflects only multiplicative effects. Everitt & Smith (1979) have confirmed that if alternative linear modeling methods based on an additive approach developed by Grizzle, Stamler and Koch (GSK) are used, interaction is present in the original Camberwell data (see also Swafford, 1980). They state that the choice between models using the two different types of interaction cannot be settled in statistical terms, but none the less show a preference for a multiplicative approach. However, recent opinion in epidemiology has emphasized how much information can be lost by ignoring additive interaction (Rothman, 1974, 1976, 1978; Blot & Day, 1978).

Elsewhere we have given what we believe are reasonably convincing pragmatic reasons for preferring an additive approach (Brown & Harris, 1986*a*: 153–154). Given the presence of such interaction the application of separate chi-square tests to the two parts of the data (as in Table 4.1) is appropriate. However, we recognize that there is no satisfactory way of finally settling on the appropriate mode of analysis by such arguments and we have therefore also presented more *theoretically* based reasons for preferring an additive model (see Brown & Harris, 1986*a*: 154–157; and Brown, 1986).

THE BASIC RESULT IN CONTINUOUS TERMS

Implicit in our presentation so far has been the assumption, noted earlier, that it is appropriate to see depression (and, for that matter, the variables involved in its onset) in categorical terms. But since we wished to explore the implications of dealing with depression in continuous terms, we developed a straightforward continuous scale of depression based on the count of PSE symptoms. But this immediately confronted us with a diagnostic decision about the particular PSE symptoms which, not being relevant for depression, should be ignored in such a procedure. We proceeded by adding to the 11 core PSE symptoms of depression presented earlier further PSE symptoms of depression, usually quite rare, together with symptoms largely dealing with tension and worry: widely used in scales such as that of the Hamilton Rating Scale (Hamilton, 1960) and the Beck Depression Inventory (Beck *et al.*, 1961). These extra symptoms were (a) inefficient thinking, morning depression, social withdrawal, lack of self-confidence, ideas of reference, guilty ideas of reference, pathological guilt, depressive delusions, delusions of reference, loss of libido, and (b) worrying, hypochondriasis, headaches (tension pains), tiredness/exhaustion, muscular tension, restlessness, nervous tension, premenstrual exacerbation and irritability.

Table 4.3 is a straightforward contingency table and shows that, once the vulnerability factor is controlled and those not at risk (because of their chronicity) are excluded, there is at best a small (and non-significant) trend for those with a higher initial symptom score to

Table 4.3. *Initial symptom score, vulnerability factor, and onset of depression at caseness level (150 Islington women with a provoking agent excluding cases at first interview)*[a]

Clinical data: continuous time 1 and categorical time 2

High vulnerability ($N = 95$)					Low vulnerability ($N = 55$)				
Initial symptom score					Initial symptom score				
0–1	2–3	4–5	6–8	9+	0–1	2–3	4–5	6–8	9+
Percentage onset case					Percentage onset case				
11	*34*	*36*	*32*	*36*	*0*	*29*	*0*	*0*	*0*
(2/19)	(10/29)	(5/14)	(6/19)	(5/14)	(0/27)	(2/7)	(0/10)	(0/8)	(0/3)
gamma = 0.23, NS					NS				

[a] Women without a provoking agent not included since only 2/153 developed onset.

Table 4.4. *Basic vulnerability result: initial symptom score, vulnerability factor, presence of provoking agent and final symptom score (303 Islington women excluding cases of depression at first interview)*

Clinical data: continuous time 1 and time 2

Overall linear multiple regression – forced entry		Beta	Significance
1.	Initial symptom score (S)	0.45	0.000*
2.	Vulnerability (V)	0.09	0.251
3.	Provoking agent (P)	0.18	0.015*
	1×2 $S \times V$	−0.02	0.830
	1×3 $S \times P$	0.02	0.811
	2×3 $V \times P$	0.12	0.160
	($R^2 = 0.42$)		

R^2 Adjusted coefficient of multiple determination.
* Significance level of at least 0.05.

experience more later onsets of caseness of depression. The implication of this result will be clear as we proceed.

Table 4.4 is crucial for our argument. It is an attempt to reproduce the basic result of Table 4.1 using multiple regression with the continuous measure of depression. It is normal practice in such analyses to include, as we have done here, the symptom score at first interview as a control factor. Material has been entered in an hierarchical (forced entry) manner in which the three main effects have been entered together with the three two-way interaction terms. The equivalent control was less complicated in the categorical analysis already shown in

Table 4.5. *Continuous score at both time 1 and time 2, initial symptom score, vulnerability factor, and onset of depression (150 Islington women with a provoking agent excluding cases of depression at first interview)*

Time 2 Symptom score	High vulnerability ($N = 95$)					Low vulnerability ($N = 55$)				
	Initial symptom score					Initial symptom score				
	0–1	2–3	4–5	6–8	9+	0–1	2–3	4–5	6–8	9+
0–1	8	3	1	2	0	11	0	2	1	0
2–3	4	10	2	2	1	9	3	1	0	0
4–5	3	1	4	0	0	4	0	3	4	1
6–8	2(1)	3	2	6	1	3	2	1	1	1
9+	2(1)	12(10)	5(5)	9(6)	12(5)	0	2(2)	3	2	1
	gamma 0.56 $\chi^2 = 16$ d.f., $p < 0.001$					gamma 0.63 $\chi^2 = 16$ d.f., $p < 0.05$				

Figures in brackets represent onset cases during follow-up – categorical variable.

Table 4.3, merely involving the omission of cases at time 1, i.e. those rated '1' on the dichotomous dependent variable of caseness of depression. Of course, with the more wide-ranging dependent variable provided by the continuous score, such a control must involve more than this. Therefore, in Table 4.4, not only have all cases of depression at first interview been excluded but the continuous score of depression at time 1 has also been entered in the regression.

Two results shown are surprising in the light of the earlier categorical analysis. First, the beta weight for the interaction between provoking agent and vulnerability factor is small and falls far short of significance. Second, the beta weight for initial symptoms far exceeds that of the only other factor of statistical significance, provoking agent: 0.45 versus 0.18.

The apparent importance of initial symptoms is, of course, in direct conflict with the earlier result using caseness which showed at best a small trend for initial continuous score to relate to caseness of depression in the follow-up period (Table 4.3). The most plausible explanation is that the continuous measure, unlike that of caseness, is essentially one of well-being rather than depression, the high beta weight for initial symptoms merely reflecting the ability of absence of symptoms at time 1 to predict the same at time 2. The beta weight is high because there is considerable variability of the score, both at time 1 and time 2, and the high resulting variances are bound to be reflected in any measure (like standardized beta coefficients) that are based on variance explained. Table 4.5, repeating Table 4.3 but using the continuous scores at both times, shows how this has happened. Not

Table 4.6. *Basic vulnerability result: initial symptom score, vulnerability factor, presence of provoking agent, and final symptom score (303 Islington women excluding cases of depression at first interview)*

Clinical data: continuous time 1 and time 2

Stepwise multiple regression entering the terms used in Table 4.4	Beta	Significance
Initial symptom score (S)	0.47	0.000*
Provoking agent (P)	0.16	0.007*
Vulnerability × provoking agent ($V \times P$)	0.19	0.002*
($R^2 = 0.431$)		

only is the degree of association between time 1 and time 2 scores markedly increased over Table 4.3, but the single most populated cell is clearly that where scores are 0–1 in both time periods ($N = 19$ for both vulnerability groups combined); in other words, one of the prime functions in Table 4.5 is the prediction of marked well-being at time 2 by marked well-being at time 1. If we return to the basic result shown in Table 4.4, it is easy to fall into the trap of believing that we see reflected the relative importance of psychiatric and psychosocial factors in determining *clinically relevant depression*, and, indeed, this appears to be the most usual interpretation in the literature of such a patterning of results (see, e.g., Warheit, 1979).

The reason why the analysis so singularly fails to reflect interaction is, however, less obvious, and we will delay discussion of this surprising result. None the less, it should be noted that when, instead of a conventional hierarchical (forced-entry) approach, stepwise linear regression is used which allows effects to emerge in order of their importance, some indication of an interactive effect does appear (Table 4.6). However, it is essential here to distinguish between an exploratory and a confirmatory analysis. It will be recalled that one of our questions is how far we would have been likely to uncover the basic vulnerability result by using a multiple regression approach. The answer is that it would appear unlikely. In the first place, many would take the position that, without a prior result indicating such an interactive effect, it would be inappropriate to utilize the kind of non-hierarchical approach resulting from this latter stepwise analysis. And it would thus have been very unlikely to have been run in the light of the negative forced entry results shown in Table 4.4, where no inkling of an interactive effect

Table 4.7. *Basic vulnerability result: initial symptom score, vulnerability factor, presence of a provoking agent and onset of caseness depression (303 Islington women excluding cases of depression at first interview)*

Clinical data: continuous time 1 and categorical time 2

Overall linear multiple regression – forced entry		
1. Initial symptom score (S)	0.15	0.175
2. Vulnerability (V)	0.04	0.659
3. Provoking agent (P)	0.03	0.729
1×2 $S \times V$	0.05	0.742
1×3 $S \times V$	−0.05	0.660
2×3 $V \times P$	0.38	0.000*
($R^2 = 0.17$)		

emerged. (It should perhaps be added that even if the initial symptom scale is omitted from the regression an interactive effect still does not emerge.) At this point therefore we conclude that the vulnerability effect clearly shown in Table 4.1 would have been missed in any exploratory analysis using multiple regression with a continuous rather than a categorical measure of depression; and that even if someone set out initially with such a result in mind they might well still have reached a negative conclusion – this time about its replicability.

Before seeking reasons for the failure of the multiple regression analysis to reveal interaction, it is important to note that even when symptoms at time 1 are measured by the same continuous score, if the dichotomous case versus non-case measure of depression is used for psychiatric status at time 2 (the dependent variable) a strong interactive effect does emerge using the forced-entry approach. Moreover the effect of initial symptom score is low and non-significant (Table 4.7). These are exactly the results that would be expected since such a binary multiple regression arrives at exactly the same estimates of main effects and interaction as the use of more traditional methods of analysing the proportions via the contingency table (Hellevik, 1984).

The problem therefore does not appear to be one of using multiple regression as such, but of using it with a *continuous* measure of depression as the dependent variable. However, using multiple regression with a dichotomous categorical outcome measure does not entirely replicate the basic vulnerability finding of Table 4.1. When borderline caseness is also included as part of a trichotomous categorical outcome variable, the beta weights of the dichotomous analysis remain practically identical. (They are so close to those in Table 4.7 that no further table has been given.) In other words we would still have missed the fact that borderline case conditions are not involved in vulnerability (Table 4.2).

Table 4.8. *Basic vulnerability result: initial symptom score, vulnerability factor, presence of a provoking agent, and final symptom score (353 Islington women including cases of depression at first interview)*

Clinical data: continuous time 1 and time 2

1. Initial symptom score (S)	0.78	0.000*
2. Vulnerability (V)	0.11	0.071
3. Provoking agent (P)	0.16	0.007*
1×2 $S \times V$	−0.15	0.073
1×3 $S \times P$	−0.05	0.561
2×3 $V \times P$	0.09	0.236
($R^2 = 0.58$)		

In terms of a wider perspective we believe it essential to be in a position to pick up such 'discontinuities', particularly when, as in this case, the result appears to be replicable (Brown & Harris, 1986*b*). As already noted, it is unnecessary in the light of the discontinuity involved in the case/borderline case/normal distinctions to assert anything about the underlying nature of the phenomenon – only that it is useful to treat the apparent continuum categorically. If the possibility of this kind of discontinuity is accepted, it will be necessary to do more than examine the coefficients of a regression analysis; the data will need to be examined specifically for such thresholds between categories in order to identify the cut-point at which the change in quantity becomes the change in quality.

One further point needs to be made about the use of continuous symptom scores. Most multiple regression studies would probably not have excluded altogether the cases of depression at the time of first interview as we did here. These would be included despite the fact that, unless they recovered early in the follow-up period, they would not be at risk for a further onset at a caseness level; but this would be justified by the general notion that in entering initial symptom score into the regression such a problem would be 'controlled for'. If the 50 who have initial caseness of depression are included in this way, findings of the earlier analysis (Table 4.4) are, if anything, exaggerated with the beta weight of initial symptom score almost doubling (0.78 versus 0.48) and again there is a failure of any interaction to emerge (Table 4.8). A re-examination of Table 4.5 on the larger population including those depressed at time 1 illuminates this. Not only is the range of scores yet further increased (several scored 21 at time 1), but the 50 cases introduce further stability to the correlation of scores between the two periods. Of the 50 only 10 had changed score by more than 2 points (8 improving and 2 worsening).

In many ways, these conclusions concerning the inclusion of those with initial caseness of depression will not surprise habitues of multiple regression. There is a whole literature describing how time-series analyses (repeated observations on the same item through time) are more likely than cross-sectional analyses (unique observations on different items at the same point in time) to give misleading importance to the variables measured more than once (Ostrom, 1978; Lewis-Beck, 1980). The low autocorrelation assumption of multiple regression requires that error for an observation at an earlier time is not related to errors for an observation at a later time and clearly this is untenable – the unmeasured forces accounting for symptom score at time 2 are not independent of the unmeasured forces influencing it at time 1.

One way of coping with this problem is to adopt a different method of controlling for initial symptoms. Instead of regressing the independent variables upon symptom score at time 2, a differential score can be calculated to indicate the degree and direction of changes in scores between the two times. When this is done for this data set, the key effect (the interaction between provoking agent and vulnerability factor) still fails to emege as statistically significant.

Therefore at this point our main conclusions are that forced-entry multiple regression, when using standardized beta coefficients, fails to produce evidence for the critical vulnerability effect; and that moreover it gives misleading weight (in the light of Table 4.3) to the relative importance of initial symptom score in predicting phenomena of clinical importance (as opposed to less substantial variations in mood). Table 4.9 summarizes material in the earlier tables in terms of initial symptom score and the vulnerability factor, although it would hold equally for provoking agent.

One possible objection to the analysis so far is our use of standardized regression coefficients to reflect causal effects and thus the relative importance of particular factors. In doing so, of course, we reflect a great deal of current practice in the social sciences where the distinction between prediction and explanation is often blurred. Achen (1982) is particularly critical of this usage. He points out how such measures (all depending on variance explained) have 'doubtful meaning but great rhetorical value': p. 59). Their value is highly dependent on the tightness of the grouping of the variables involved and this by no means reflects the 'strength' of the underlying relationship. The example he discusses illustrates how the R^2 of two regressions coefficients can be totally at odds with an actual measure of 'level of importance' based on the unstandardized coefficients. He points out that such an index of relative importance has the attractive property that, when the contributions of the independent variable to be contrasted are added, together with the intercept, the result is precisely the mean of the dependent variable (Achen, 1982: 72).

Table 4.9. *Relative size beta weights of initial symptom score and vulnerability factor in terms of various outcome measures (150 Islington women with a provoking agent)*

Clinical data: various

	A. Outcome: (categorical)		B. Outcome: caseness and borderline caseness (categorical)		C. Outcome: symptom score (continuous)	
	Beta	Sig.	Beta	Sig.	Beta	Sig.
1. Initial symptom score (150 women excluding cases at first interview)	0.10	0.206	0.14	0.092	0.45	0.000*
2. Vulnerability	0.29	0.000*	0.28	0.001*	0.18	0.013*
Ratios 1/2		0.35		0.48		2.49

	D. outcome: symptom score (continuous)	
	Beta	Sig.
1. Initial symptom score (189 women including cases at first interview)	0.64	0.000*
2. Vulnerability	0.11	0.047*
Ratio 1/2		5.82

As it happens, this does not appear to explain the difference in relative importance of the clinical and social measures in the results so far presented. For example, in Table 4.4 the ratio of the two beta coefficients of initial symptom score and provoking agent is 2.04, whereas if Achen's level of importance measure is used this ratio is still much the same at 2.24.

There are other complexities which we will not pursue (e.g., whether non-significant interaction factors should be ignored). However, we trust we have conveyed that the issue of the relative importance of various factors needs to be argued carefully. Given the point made earlier about the continuous clinical measures reflecting well-being, multiple regression as a technique appears to place misleading weight in the present instance on clinical as against social phenomena.

So far we have placed a good deal of weight upon more mathematical considerations, as a possible explanation of the discrepant findings about the presence of an interaction effect. However we also wondered whether there might not be more substantive issues of measurement involved. For example, could another continuous score better capture the essence of depression?

OTHER CONTINUOUS MEASURES OF DEPRESSION

So far for the continuous depression score we have used quite a *general* measure because it appears to reflect other current measures reasonably well. The score ranged from 0 to 13 PSE symptoms at time 1 (excluding cases at first interview) with a mean of 2.78. But symptoms such as tension and lack of energy can exist in the absence of a core depressive condition and could therefore represent 'noise' as far as a measure of depression is concerned. We therefore examined the consequences of using a continuous measure that followed more closely our own definition of depression, taking account of depressed mood and only the 10 other symptoms of depression listed when the categorical measure was outlined earlier. (This shortened measure of core symptoms will be referred to as symptom score B.) The score at time 1 now only ranged from 0 to 7 (excluding cases at first interview) with a mean of 0.67.

With this less dispersed measure the interactive effect of provoking agent and vulnerability factor emerges clearly with a forced-entry multiple regression (Table 4.10). The beta coefficient for initial symptoms is of much the same order as before (Tables 4.10 and 4.4). None the less in the light of our earlier argument it is still doubtful whether the relative size of the beta weights for initial symptom score (0.53) and interaction between provoking agent and vulnerability factor (0.37) should be accepted as reflecting their relative importance. Interestingly enough when the cases at first interview are included, although the beta coefficient for initial symptoms is particularly high, the interaction between provoking agent and vulnerability factor is still the second most important effect (Table 4.11).

Thus conclusions reached with score B are radically different from those with the first continuous measure. Results follow much more closely the original findings based on the categorical measure of caseness. The best-fitting model would include the interaction between provoking agent and vulnerability. There is one further consistency: if a

Table 4.10. *Basic vulnerability result: initial symptom score B, vulnerability factor, presence of a provoking agent, and final symptom score B (303 Islington women excluding cases of depression at first interview)*

Clinical data: coninuous time 1 and time 2

Overall linear multiple regression – forced entry	Beta	Significance
1. Initial symptom score B (S)	0.53	0.000*
2. Vulnerability (V)	−0.01	0.875
3. Provoking agent (P)	0.09	0.237
1×2 $S \times V$	−0.09	0.387
1×3 $S \times P$	−0.18	0.100
2×3 $V \times P$	0.37	0.000*
($R^2 = 0.303$)		

Table 4.11. *Basic vulnerability result: initial symptom score B, vulnerability factor, presence of a provoking agent, and final symptom score B (353 Islington women including cases of depression at first interview)*

Clinical data: continuous time 1 and time 2

Overall linear multiple regression – forced entry	Beta	Significance
1. Initial symptom score B (S)	0.91	0.000*
2. Vulnerability (V)	0.04	0.561
3. Provoking agent (P)	0.08	0.195
1×2 $S \times V$	−0.36	0.001*
1×3 $S \times P$	−0.12	0.271
2×3 $V \times P$	0.27	0.001*
($R^2 = 0.46$)		

tabular approach is used for initial symptoms (as earlier in Table 4.3) there is now a better case to be made that they are of some significance (Table 4.12). The relationship of initial symptom score and caseness of depression now reaches statistical significance. (We had in fact in an earlier analysis recognized the importance of a chronic borderline case condition of either depression *or* anxiety as a factor predicting subsequent onset of case depression and explored this using logistic regression: see Brown *et al.*, 1986).

The reasons for these dramatic changes, following the changes in the scale used, are by no means clear. It seems likely that the longer

Table 4.12. *Initial symptom score B, vulnerability factor, and onset of depression at caseness level (150 Islington women with a provoking agent excluding cases of depression at first interview)*

Clinical data: continuous time 1 and categorical time 2

High vulnerability ($N = 95$)					Low vulnerability ($N = 55$)				
Initial symptom score B					Initial symptom score B				
0	1	2	3	4+	0	1	2	3	4+
Percentage onset case					Percentage onset case				
15 (6/40)	*57* (12/21)	*31* (4/13)	*20* (2/10)	*36* (4/11)	*3* (1/34)	*10* (1/10)	*0* (0/8)	*0* (0/2)	*0* (0/1)
		40 (22/55) gamma = 0.31, $p < 0.02$, 4 d.f.					*5* (1/21) NS		

Table 4.13. *Basic vulnerability result: initial ID (Index of Definition), vulnerability factor, presence of a provoking agent and final ID (303 Islington women excluding cases of depression at first interview)*

Clinical data: continuous time 1 and 2

Overall linear multiple regression – forced entry	Beta	Significance
1. Initial symptom ID (S)	0.44	0.000
2. Vulnerability (V)	0.08	0.538
3. Provoking agent (P)	0.34	0.006
1×2 $S \times V$	0.06	0.679
1×3 $S \times P$	−0.15	0.307
2×3 $V \times P$	0.06	0.539
($R^2 = 0.36$)		

measure had greater variability owing to the inclusion of non-depressive symptoms and this in some manner swamped the interactive effect of provoking agent and vulnerability factor. In order to investigate this we looked at a third continuous measure of symptoms – the Index of Definition of the PSE (or ID) – which has been used in multiple regression analyses (see, e.g., Tennant, Bebbington & Hurry, 1982). This is an algorithm developed for the PSE to reflect the chances of a symptom picture reflecting a psychiatric case of any kind, not only of depression; it has a more modest range of scale points than the total PSE score which has been usually employed, running from 1 (no disorder) to 8 (most definite caseness). Table 4.13 shows that with this third measure the interactive effect does not emerge as significant in a

Table 4.14. *Initial symptoms with score B by Index of Definition: women in Islington followed up, excluding cases of depression (N = 303)*

		Score B (Low)						
		0	1	2	3	4	5–7	Total
	(Low)							
(Index of definition)	1	*55*						55
	2	83	*15*	5				103
	3	38	24	*10*	1			73
	4	8	10	11	*7*	0	1	37
	5	5	2	8	6	*5*	5[a]	31
	6	0	0	0	0	1	*3*[b]	4
Total		189	51	34	14	6	9	

[a] on score B 5 = 3 and 6 = 2.
[b] on score B 5 = 1, 6 = 1 and 7 = 1.

multiple regression despite the apparent similarity of ID to score B in terms of distribution.

At time 1 the distribution of ID is: 1 & 2:158, 3:73, 4:37, 5:31, and 6:4; and for score B: 0:189, 1:51, 2:34, 3:14, 4:6, 5:5, 6:3 and 7:1 (see marginal totals in Table 4.14). A further look at the two scores, however, in Table 4.14 suggests that they are picking up different phenomena: there were many more scoring low on score B and high on ID than high on Score B and low on ID. This is not surprising, given our knowledge that the ID level picks up conditions other than depression while score B is designed to mimic the Bedford College depressive caseness criterion. It suggests that the non-emergence of the interactive effect between provoking agent and vulnerability may result from the introduction of these other conditions into the continuous scale which therefore fails to reflect a dimension of pure depression. This introduction of 'noise' may be even more important in explaining the failure of the original continuous score to reflect interaction (Table 4.4) than our earlier speculations concerning the fact it tended to reflect well-being with a consequent relative neglect of depressive disorder as such.

SUMMARY AND CONCLUSIONS

In the light of the basic findings using categorical measures of depression (Tables 4.1 and 4.2) we find it difficult to give much weight

to the results of the initial forced entry multiple regression analysis using a continuous measure of depression (Table 4.4). The most likely reason for the failure to find an interactive effect between provoking agent and vulnerability factor in predicting subsequent depression is that the general and continuous symptom measure includes too much that is strictly unrelated to depression. There is already evidence that the vulnerability model involving provoking agent and support does not hold for the onset of anxiety conditions (Finlay-Jones, 1989), and for this reason we had deliberately omitted a count of anxiety symptoms from this larger continuous measure. But it will also follow that the effect is likely to be attenuated by the inclusion of the tension conditions that form the bulk of other diagnoses (Brown, Craig & Harris, 1985). However, it is also possible that the lack of an interactive effect relates to the greater variability inherent in the general scales and this in some way influences the standardized coefficients that reflect variance. It is of interest here to note the result of using the third psychiatric measure, the Index of Definition (ID), which, although resembling score B in variability, did not identify the key interactive effect as significant, thus resembling the longer score. This suggests that the failure of the general scales may be more related to the nature of the symptoms they combine, which introduces noise into the dimension of pure depression, rather than to the distribution of the scores as such.

A second reason to doubt the initial forced-entry regression analysis is that the preeminent importance given to the continuous initial symptom score compared with that of the established social aetiological factors (Table 4.4) runs quite counter to the, at best modest and non-significant, association between initial symptom score and onset of *caseness* of depression (Table 4.3). Here the high beta weight almost certainly reflects the ability to predict well-being at one point in time on the basis of well-being at an earlier point. In short it would be misleading to see such a measure as crucially focusing upon *clinically* relevant psychiatric phenomena.

We have therefore been forced to the conclusion that the use of multiple regression as the sole technique with the original long clinical measure would have failed to reveal the basic interactive effect necessary for the vulnerability result and may well have led to quite erroneous conclusions about the relative importance of subclinical conditions and social factors in predicting future depression. This most clearly holds for any study not setting out with a particular hypothesis about the vulnerability effect. However, even if multiple regression had been used in a study specifically designed to replicate the earlier findings, it might well have led to the conclusion that there was no replication. It is true that a stepwise approach to regression would have been able to reveal some evidence for the relevant interaction effect (Table 4.6); however,

this might well not have been carried out, given the negative results of the forced entry, and some would question the status of such a non-hierarchical approach when the initial findings, involving the 'main effects', were negative.

Finally, the common practice of including all depressive conditions (whether or not they had at first interview reached caseness of depression in our terms) would only have compounded these problems (Table 4.8).

Despite these difficulties, when the long version of the continuous clinical measure was used as a control measure (for state at time 1) and a categorical measure of caseness used as the outcome measure, the expected interactive effect clearly emerged as the only effect of statistical significance: i.e., initial symptom score did not relate to outcome (Table 4.7). This, as already noted, is expected from the fact that multiple regression when used with a dichotomous dependent variation parallels the results of more traditional analyses in terms of differences in proportions (Hellevik, 1984). However, even this result has the shortcoming that it cannot reveal that onsets of borderline case conditions do not reflect a vulnerability effect (Table 4.2).

This initial largely negative exercise with multiple regression led us to reconsider our continuous measure of depression (which included strictly non-depressive symptoms such as worry, tiredness and muscular tension) as it seemed unlikely that the negative results could be solely attributed to the use of multiple regression as such. The use of a shortened measure of depression (score B) in fact produced a radically different set of findings. The expected interactive effect clearly emerged (Table 4.10). The trouble is that there is no reason to believe that such an instrument would be the measure of choice for those wedded to a regression approach. It needs to be borne in mind that the shortened scale, using only 11 core symptoms of depression, was derived from our 'categorical' measure of depression. Moreover, the results dealing with the shortened version still leave obscure the question of the relative importance of initial symptoms and social factors in predicting subsequent depression. We believe it therefore would be highly misleading to take literally the relative importance of the relevant beta weights when using score B (Table 4.10).

The issues here are complex, and perhaps it is sufficient to state our belief that the contribution of symptoms falling short of caseness is best explored by traditional tabular analyses and this followed up by logistic regression using categorical measures of clinical state. For example, a straightforward tabulation makes clear that the predictive importance of symptoms at time 1 (once cases of depression are excluded) appears to involve all diagnostic conditions, particularly of anxiety, as long as they are chronic (Brown *et al.*, 1986: Tables 2 and 3). However, their importance appears to be related to their correlation with ongoing difficulties: chronic subclinical symptoms present at the time of first interview are entirely unrelated to subsequent development of caseness depression once such background difficulties and the severe events (to which they often lead) are taken into account (Brown *et al.*, 1986). In other words we are not rejecting multivariate techniques as such – but advocating their use in parallel with more traditional methods.

It is perhaps in order here to cite just one more result since it further underlines the importance of a categorical approach. In an exercise, not so far reported, the original continuous measure (which did not produce an interactive effect) was used as a dichotomy. The cut-point used defined 21% of the sample as 'cases' of depression in the follow-up year (63/303) once the chronic Bedford College cases were excluded. This did produce a statistically significant interactive effect for provoking agent and vulnerability factor unlike its counterpart analysis with the continuous measure in Table 4.4. This measure of 'caseness' as well as including all 32 onsets of depression included 21 of the 27 onsets of borderline depression. Given the inclusion of the latter (and 10 without an onset of a depressive condition), it would be expected that the beta weight for the interaction between provoking agent and vulnerability factor is somewhat lower than with symptom score B (0.26 versus 0.37: see Table 4.10). The important point is that a measure of depression apparently unsuitable for use as a continuous measure produced the 'correct' result when used as a dichtomy.

In the light of these comments it is interesting to reconsider some of the conclusions often cited in the literature. For example, Akiskal cites two studies as evidence that psychosocial factors 'seem to determine the timing of depressive onsets, rather than their fundamental causes' because the authors found that 'the variance contributed by life stressors was quite modest, compared with that contributed by a personal history of depression' (Akiskal, 1985: 133). One of these studies had carried out an analysis not unlike our Tables 4.8 and 4.11 here, but over a three-year rather than a one-year follow-up period, concluding that time 1 depression scores accounted for more explained variance (both unique and shared) than the other factors, losses (like our provoking agents) and resources (like our vulnerability) (Warheit, 1979). Although most of the 18 items of the depression scale used are relatively specific to depression rather than tension or anxiety, there was still scope for noise to enter the dimension through some of the less specific items such as tiredness and somatic complaints. It is therefore difficult to tell whether the depression score used by Warheit would more resemble the longer one used in Table 4.8 or score B used in Table 4.11 but, as we have argued, his inclusion of 'cases' at time 1 would be bound to have raised the beta coefficient of initial symptom score, and to draw conclusions

about the relative causal importance of factors on that basis would be misleading in the way we have outlined. The second study spanned a nine-year follow-up period and examined its series of multiple logistic models twice, once including and once excluding those who were cases at first interview (Kaplan, Roberts, Camacho & Coyne, 1987). The authors comment that it is 'perhaps not surprising' that the strongest predictor of a high level of depressive symptoms in 1974 was a high level in 1965, and in general do not adopt the rather superior attitude towards the variable embodying the psychosocial environment which Akiskal seems to attribute to them. They conclude 'what is most important is that the pattern of depressive symptoms at baseline status . . . psychosocial factors and physical health problems are major predictors of high levels of depressive symptoms'. They have not, however, given models with interaction terms.

Another analysis often cited in the literature used multiple regression to explore the effects of a number of childhood and adult demographic variables upon adult psychiatric morbidity, measured by the Index of Definition (Tennant *et al.*, 1982). This was a cross-sectional analysis, so the problems arising from the inclusion of initial symptom score, or cases at first interview, do not arise; but another issue raised in our analysis might have some relevance and so is worth airing. This concerns the use of ID rather than a purer measure of depression. The authors themselves address the possibility that the estimate of variance explained by childhood loss and deprivation might have been higher if there had not been, what we have here called 'noise', introduced by their choice of dependent variable, and mention that 'most subjects with "disorder" had depressive syndromes (67%)' (p. 326). However the noise which concerns them is rather different from the noise we focused on here in discussing ID. The very fact that they slip so easily into talking in terms of a dichotomous category (with or without 'disorder') suggests that they are forgetting that the dispersion of ID points 1 to 4 is going to prove just as important for their analysis as the points above the usual threshold taken for caseness – level 5 – which is presumably where they divide those with and without disorder. For us the noise is introduced by the respondents with zero on score B but scores of 2 and above on ID (see Table 4.14). Tennant and colleagues, however, are concerned about the possibility that respondents with 'transient distress responses' might be diluting the true depressions, and suggest that the way to distinguish these would be to determine which were caused by life events rather than to incorporate further symptom criteria to make the distinction. Nevertheless, despite our substantially differing definitions of the noise, Tennant and colleagues are highlighting a similar general issue to the one identified in this chapter: that estimates of a variable's importance occasionally need an upward revision if the variable concerned really specifically affects depression, and the measure used has been the ID.

What is interesting is that all of the three studies just cited have felt the need to utilize categorical measures at some point in the reports, usually in describing how many were depressed or suffered from disorder at a given stage. But then they have gone on to analyse the data using a continuous measure. Finally we should not omit to mention a fourth study which, with concepts similar to those discussed here, did analyze the data twice: once treating depression as an interval variable (CES-D), and once as a dichotomy – with the cut-off at 16 and above (Lin & Ensel, 1984). In neither analysis did they confirm any effect of the interaction between life events (corresponding to our provoking agent) and social support (corresponding to our vulnerability), but this may be the result of their using *change* in CES-D score rather than CES-D score at time 2, and especially change in life event score and change in social support, as the key variables. As will be appreciated, the effect of social support upon depressive onset is not necessarily through a *change* in its amount but more usually through an absolute low level which may be unchanging. Lin & Ensel control for initial symptom score by entering CES-D at time 1 (CES-D1) along with change in life events and change in social support, and find it has a higher beta weight than the latter two, but do not then move on to attribute greater aetiological importance to CES-D1 than to psychosocial factors. This is possibly because the beta weight here is negative as a result of the nature of the dependent variable (change in CES-D score rather than its absolute level at time 2). However another series of analyses of the same data focusing on CES-D2 (Lin & Dean, 1984) does not automatically enter CES-D1 along with the psychosocial variables: this suggests that in general they are considerably less preoccupied with this issue of the relative importance of symptomatological and psychosocial factors than the other authors cited earlier.

To sum up then, we have identified a set of possible analyses which, if reported, could prove genuinely misleading. The origins of these misconceptions may lie as much in substantive issues of measurement as in the use of the regression procedure itself. In other words, the failure to appreciate discontinuities in the aetiological process may arise more from deficiencies in conceptualizing the specific nature of depression as opposed to psychiatric disorder in general, than from the use of continuous scores and the consequent failure to employ categories with boundaries which coincide with this discontinuity. After all, it could be said that specificity is conceptually allied to discontinuity. Yet it could be argued that it is exactly the beguiling accessibility of computerized multiple regression programmes requiring a continuous score for the dependent variable that tempts research workers into the use of

dimensions which do not embody the phenomenon they set out to study.

Finally, it has become much clearer to us that the shortcomings of multiple regression in the present instance primarily involve its use in opening up a field of enquiry. It would clearly be possible with the exercise of some ingenuity to replicate and make sense of the basic findings concerning the vulnerability effect in terms of multiple regression as long as we are willing to deal with effects non-hierarchically. If we are forced to draw a conclusion, other than underlining the possible pitfalls of multiple regression as a technique of choice for studying clinically relevant phenomena, it is to emphasize the wisdom of employing a number of approaches in opening up an area of enquiry, and particularly that multivariate statistical techniques should not be seen as replacing traditional methods involving direct 'eyeballing' of data.

ACKNOWLEDGEMENTS

The Islington study was funded by the Medical Research Council London. We are grateful to Laurie Letchford for his tireless help with the computer, to Sue Ross for preparation of the manuscript and to Martin Eales and Andrew Pickles for their comments on an earlier draft.

REFERENCES

Achen, C. H. (1982). Interpreting and using regression. Sage University Paper series on: *Quantitative applications in the Social Sciences*, (eds.) John L. Sullivan & Richard G. Niemi. Sage Pubns.

Akiskal, H. S. (1985). Interaction of biologic and psychologic factors in the origin of depressive disorders. *Acta Psychiatrica Scandinavica, 71,* 131–139.

Bebbington, P. (1980). Causal models and logical inference in epidemiological psychiatry. *British Journal of Psychiatry, 136,* 317–325.

Beck, A. J., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

Blot, W. J. & Day, N. E. (1978). Synergism and interaction: are they equivalent? *American Journal of Epidemiology, 110,* 99–100.

Brown, G. W. (1986). Statistical interaction and the role of social factors in the aetiology of clinical depression. *Sociology, 20,* 135–140.

Brown, G. W. (1989*a*). Causal paths, chains and strands. In M. Rutter (ed.), *The power of longitudinal data: Studies of risk and protective factors for psychosocial disorders.* Cambridge University Press.

Brown, G. W., Bifulco, A. & Andrews, B. (1990). Self-esteem and depression III. Aetiological issues. *Social Psychiatry and Psychiatric Epidemiology, 25,* 235–243.

Brown, G. W. & Harris, T. O. (1978*a*). *Social origins of depression: A study of psychiatric disorder in women.* London: Tavistock Publications.

Brown, G. W. & Harris, T. O. (1978*b*). Social origins of depression: A reply. *Psychological Medicine, 8,* 577–588.

Brown, G. W. & Harris, T. O. (1980). Further comments on the vulnerability model. *British Journal of Psychiatry, 137,* 584–585.

Brown, G. W. & Harris, T. O. (1986*a*). Establishing causal links: The Bedford College studies of depression. In H. Katschnig (ed.), *Life events and psychiatric disorders* (pp. 107–187). Cambridge University Press.

Brown, G. W. & Harris, T. O. (1986*b*). Stressor, vulnerability and depression: A question of replication. *Psychological Medicine, 16,* 739–744.

Brown, G. W. & Prudo, R. (1981). Psychiatric disorder in a rural and an urban population. 1. Aetiology of depression. *Psychological Medicine, 11,* 581–599.

Brown, G. W., Craig, T. K. J. & Harris, T. O. (1985). Depression: Distress or disease? Some epidemiological considerations. *British Journal of Psychiatry, 147,* 612–622.

Brown, G. W., Bifulco, A., Harris, T. O. & Bridge, L. (1986). Life stress, chronic subclinical symptoms and vulnerability to clinical depression. *Journal of Affective Disorders, 11,* 1–19.

Brown, G. W., Bifulco, A. & Harris, T. O. (1987). Life events, vulnerability and onset of depression: Some refinements. *British Journal of Psychiatry, 150,* 30–42.

Cooper, J. E., Copeland, J. R. N., Brown, G. W., Harris, T. O. & Gourlay, A. J. (1977). Further studies on interviewer training and inter-rater reliability of the Present State Examination (PSE). *Psychological Medicine, 7,* 517–523.

Endicott, J. & Spitzer, R. L. (1978). A diagnostic interview – The schedule for affective disorders and schizophrenia. *Archives of General Psychiatry, 35,* 837–844.

Everitt, B. S. & Smith, A. M. R. (1979). Interactions in contingency tables: a brief discussion of alternative definitions. *Psychological Medicine, 9,* 581–583.

Finlay-Jones, R. (1989). Anxiety. In G. W. Brown & T. O. Harris (eds.) *Life events and illness.* New York: Guilford Press. London: Unwin Hyman.

Finlay-Jones, R. & Brown, G. W. (1981). Types of stressful life events and the onset of anxiety and depressive disorders. *Psychological Medicine, 11,* 803–815.

Finlay-Jones, R., Brown, G. W., Duncan-Jones, P., Harris, T. O., Murphy, E. & Prudo, R. (1980). Depression and anxiety in the community: Replicating the diagnosis of a case. *Psychological Medicine, 10,* 445–454.

Goldberg, D., Cooper, B., Eastwood, M., Kedward, H. & Shepherd, M. (1970). A standardized psychiatric interview for use in community surveys. *British Journal of Preventive and Social Medicine, 24,* 18–23.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology and Neurosurgical Psychiatry, 23,* 56–62.

Hellevik, O. (1984). *Introduction to causal analysis.* London: Allen and Unwin.

Kaplan, G. A., Roberts, R. E., Camacho, T. C. & Coyne, J. C. (1987). Psychosocial predictors of depression. Prospective evidence from the human population laboratory studies. *American Journal of Epidemiology, 125*(2), 206–220.

Lewis-Beck, M. S. (1980). Applied regression – an introduction. Sage University Paper series: *Quantitative applications in the social sciences.* Beverly Hills and London: Sage Pubns.

Lin, N. & Dean, A. (1984). Social support and depression: A panel study. *Social Psychiatry, 19,* 83–91.

Lin, N. & Ensel, W. (1984). Depression-mobility and its social etiology: The role of life events and social support. *Journal of Health and Social Behaviour, 25,* 176–188.

Medawar, P. B. (1969). *Induction and intuition in scientific thought.* London: Methuen.

Ostrom, C. W., Jr (1978). Time series analysis: Regression techniques. Sage University Paper series: *Quantitative applications in the social sciences.* Beverly Hills and London: Sage Pubns.

Prudo, R., Brown, G. W., Harris, T. O. & Dowland, J. (1981). Psychiatric disorder in a rural and an urban population: 2. Sensitivity to loss. *Psychological Medicine, 11,* 601–616.

Rothman, K. J. (1974). Synergy and antagonism in cause-effect relationships. *American Journal of Epidemiology, 99,* 385–388.

Rothman, K. J. (1976). The estimation of synergy or antagonism. *American Journal of Epidemiology, 103,* 506–511.

Rothman, K. J. (1978). Occam's razor pares the choice among statistical models. *American Journal of Epidemiology, 108,* 347–349.

Schroeder, L. D., Sjoquist, D. L. & Stephan, P. E. (1986). Understanding regression analysis – an introductory guide. Sage University Paper series: *Quantitative applications in the social sciences.* Beverly Hills and London: Sage Pubns.

Swafford, M. (1980). Three parametric techniques for contingency table analysis: a nontechnical commentary. *American Sociological Review, 45,* 664–690.

Tennant, C. & Bebbington, P. (1978). The social causation of depression: A critique of the work of Brown and his colleagues. *Psychological Medicine, 8,* 565–575.

Tennant, C., Bebbington, P. & Hurry, J. (1982). Social experiences in childhood and adult psychiatric morbidity: A multiple regression analysis. *Psychological Medicine, 12,* 321–327.

Warheit, G. J. (1979). Life events, coping, stress, and depressive symptomatology. *American Journal of Psychiatry, 136,* 502–507.

Wing, J. K., Cooper, J. E. & Sartorius, N. (1974). *The measurement and classification of psychiatric symptoms: an instruction manual for the Present State Examination and CATEGO Programme.* Cambridge University Press.

Wing, J. K., Henderson, A. S. & Winckle, M. (1977) The rating of symptoms by a psychiatrist and a non-psychiatrist: A study of patients referred from general practice. *Psychological Medicine, 7,* 713–715.

5 Differential health in a

LEIV S. BAKKETEIG,
SUNDET

The area of research suggestec analytical epidemiology. The demonstrate in a valid and re individuals or groups of indivi time, and to describe alteratio of time. This makes it possil longitudinal health patterns. explain this variation in terms The focus in this presentation analytic epidemiology. Some cussed, but the main emphasi employing data from a pop Norway.

The analytical designs can b cross-sectional studies, and lo individuals or groups of indivi

Let us first briefly look a ecological approaches have be ment of health in a life-span p for being used as an indicator statistics or other routinely cc ment of mortality (be it crude over time can be of interest to areas or populations.

In Norway, in the middle Sundt, did some highly inter mortality in the country using the geographical areas. 130 y Aase, in Trondheim has re mortality in different regions over this long period (Aase namely parts of western Norv