Part I

The current model for depressive disorders and its impact on clinical management

1

A 'Declaration of Independence'

This book is provocative, deliberately so. It reflects frustration and disappointment at current modelling and consequently current treatments of depressive disorders. We will argue that conceptualising and modelling the depressive disorders along a dimensional continuum of severity has led to sterility of thought, clinical practice and research findings. It is time to make a 'Declaration of Independence'.

To paraphrase an address to the US Congress in 1776:

> When in the course of human events, it becomes necessary to dissolve the dimensional view spuriously connecting one depressive condition with another, and to assume that separate or relatively independent conditions exist, a decent respect to the opinions of mankind requires that we should declare cause which impel us to argue for such separation.
>
> We hold the following truths to be self-evident – that the depressive disorders are not all equal and that those endowed by their creator to enjoy the unalienable rights of life, liberty and the pursuit of happiness, are also entitled to a more sophisticated assessment and management model if they develop a depressive condition. We further declare that, as the current model of depression is destructive to that objective, it is the right of the people to expose its limitations, criticise it, and pour scorn on its quality of explanatory power. Prudence, indeed, dictates that DSM and ICD models long established should not be changed for light and transient causes but, all experience has shown, organisations are more disposed to continue, despite the associated sufferable nonsense, than to right themselves by abolishing the forms to which they have become accustomed. However, when a long train of conceptual poverty and management failures occur, it is our right, and it is our duty, to throw off such strictures and to provide new models for the individual patients and for the satisfaction of the therapist. We therefore appeal to readers to be absolved from all allegiance to an all-explanatory dimensional model, whether ideologically or politically developed, and to a greater confidence in and firmer reliance on a model allowing depressive conditions to be viewed as possessing relative independence from each other.

Having got the polemic off our chest, we will now proceed to argue key points in a manner somewhat more consistent with the academic tone of Cambridge University Press monographs, although every now and again we may move to a 'smoking' zone.

There has been an increasing tendency over the last few decades to view 'depression' as an 'it'. 'It' is then frequently interpreted according to the professional's particular paradigm. For instance, psychiatrists increasingly view 'it' as a disease reflecting a biochemical process and therefore likely to benefit from physical treatments such as antidepressant medication. Psychologists judge it to reflect schematic or attributional errors by the individual in viewing themselves, the world and their future, and therefore requiring cognitive behaviour therapy (CBT). To counsellors, 'it' reflects a disjunction between the individual and a range of social problems, and therefore benefiting from counselling and problem solving. The blind man's definition of the elephant springs to mind here. In practice, we currently have a situation where the professional background, training or disciplinary interests of the practitioner so shape the view of depression and its management that the depressed patient tends to be 'fitted' to the therapy rather than the therapy being shaped to respect nuances of the patient's particular depressive condition.

How has such confounding occurred? Let us document a few stages and a few markers along the road to obfuscation.

The wrong model

The current adherence to a dimensional model of 'depression', with conditions (if distinguished) differentiated largely on the basis of severity, has imposed a major limitation. Since the introduction of the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III) (American Psychiatric Association, 1980), the North American model has viewed the depressive disorders as subdivided into 'major' and 'minor' disorders, and this division was extended by the formulation of less severe (e.g. 'sub-syndromal' and 'sub-clinical') expressions. Logically, how can a 'sub-clinical' expression have clinical status? The World Health Organisation's latest International Classification of Diseases, 10th edition (i.e. ICD-10) (World Health Organization, 1992) is also underpinned by a severity-based model, whereby depressive disorders are subdivided into 'severe', 'moderate', and 'mild'. Such models are variants of the former 'unitarian' model, which viewed depression as a single condition varying by severity. Both the DSM and ICD systems, largely by rejecting phenomenological definition and ignoring aetiology, use severity as the definitional marker, with the DSM system also dimensionalising duration and recurrence parameters.

As in any other medical field, any attempt to then create categories from dimensionally based data, risks producing pseudo-entities or pseudo-categories. This limitation is best exemplified in the DSM concept of 'major

depression' (or 'clinical depression') as having diagnostic specificity and therefore being informative. As noted in our earlier monograph (Parker & Hadzi-Pavlovic, 1996), such a model has not generated replicable biological changes or correlates at a satisfactory level, and has not been informative in identifying treatment-specificity effects.

The last point is under-appreciated. Imagine for the moment that you are not a health professional reading this book, but an intelligent consumer who, after examining a list of DSM disorder criteria, realised that you had 'major depression'. You might wish to find out what might best 'work' for your condition. You might further want to know the evidence base for available treatments. Your reading is reassuring, at first. You are encouraged to find that psychiatry is no longer a sect-weighted field, but a *scientific discipline*, respecting an evidence base, and *weighting information* that comes from randomised control trial (RCT) studies – both evaluating treatments in comparison to each other and in comparison to placebo or control interventions. You are further reassured to learn that the database of RCT efficacy studies is the largest database existing in psychiatry, involving hundreds of thousands of subjects, and therefore capable of precise interpretation. You find that study parameters of improvement are standardised (e.g. a 'responder' is someone who has improved 50% or more over the trial period), so allowing differing studies to be compared against each other. In fact, you learn that such databases are so informative that they allow numerous explicit and authoritative sets of treatment guidelines to be generated in regard to the management of major depression.

The trouble with RCTs

But, when you look closer, you see empiricism without clothes. As noted in recent publications (Parker et al., 2003a; Parker, 2004), in reality the database is not particularly meaningful. Examples:

- One analysis of 'old' antidepressants (such as tricyclics) and new antidepressants (such as selective serotonin reuptake inhibitors, SSRIs) involving 150 studies and over 160,000 subjects returned a responder rate of 54% for each class (Williams et al., 2000).
- A more specific meta-analysis of 102 trials comparing only tricyclics and SSRIs found no difference in efficacy rates across these two classes (Anderson, 2000).
- A meta-analysis of psychotherapy trials compared to pharmacotherapy found only trivial superiority to pharmacotherapy (Robinson et al., 1990).

- Meta-analyses of nearly 30 RCTs comparing CBT, interpersonal psychotherapy and behaviour therapy reported responder rates of 50%, 52%, and 55%, respectively, suggesting comparable efficacy (DHHS Depression Guideline Panel, 1993).

In fact, when we look at responder rates in studies of antidepressant medication, psychotherapy, counselling, herbal treatment, and even bibliotherapy (i.e. reading books on depression), the responder rates are in the 50–55% range, suggesting comparable efficacy for all such strategies. Such non-specific results risk the cynical interpretation that All Roads Lead to Rome – or, at least that all treatments are equally potent in terms of managing 'major depression'. As noted by Holmes (2002), such results deliver the Dodo Bird verdict ('Everyone has won, and all must have prizes').

Such results could challenge our faith in RCTs and, as with religious agnosticism, there is the risk of abandoning religion merely because we have a problem with the local minister. The question here is whether the weighting to RCT-generated evidence is the obfuscating factor, or the procedures used in undertaking RCTs. We argue for the latter.

If we examine the RCT procedures for antidepressant drugs, the processes seem reasonable, on the surface. The study design procedures (and their voluminous documentation) appear rigorous, while their undertaking is presided over by august bodies such as the US Food and Drug Administration (FDA) who impose numerous procedures and strictures to ensure standardisation of measurement, evaluation, and the handling of those who withdraw or drop out. There is a sense of rigorous and hard science. The devil lies less in the detail however, and more in the aggregate because, if we keep our eye on the whole, then counter-intuitive findings can be explained.

A key finding is that if we examine the trajectory of improvement reported in FDA-supported studies of antidepressants compared to placebos, the difference in improvement rates is slight, at best. Two published analyses support that interpretation. In one study (Kirsch et al., 2002), the researchers examined the randomised control data for the six antidepressants approved by the FDA over the period 1987–1999 (i.e. citalopram, fluoxetine, nefazodone, paroxetine, sertraline, and venlafaxine). In the 47 trial data sets submitted to the FDA, there was no differential efficacy between the antidepressant and the placebo in nine studies while, for the remainder, the drug-placebo difference was two points on the Hamilton scale (Hamilton, 1960), which was interpreted by the researchers as 'very small and of questionable significance'. In another analysis of 52 pivotal FDA placebo-controlled studies of antidepressants (Khan et al., 2002), antidepressant efficacy was indistinguishable from placebo in 50% of the studies.

You may suggest that such results reflect a methodological limitation of meta-analysis and that, as the whole does not truly summate the parts, results would be different if individual studies were examined. However, if we turn to individual studies, similar counter-intuitive findings are evident. One is noteworthy – a study examining the efficacy of St John's Wort in comparison to both an SSRI and to placebo (Hypericum Depression Trial Study Group, 2002). Noteworthy, because the study design was impeccable (in terms of meeting, if not exceeding, standard FDA requirements). The sample was large (with a minimum cell size of 111 subjects), the study period (of 8 weeks) was sufficiently lengthy to address the particular research question, the study was overseen by several distinguished institutions, and the researchers comprised many of the most respected psychiatrists in North America. But the results? At the end of 8 weeks, the reduction in Hamilton depression scores was 27% for St John's Wort, 28% for placebo, and 29% for the SSRI, differences that are clearly trivial. The study is noteworthy in that its results are again counter-intuitive to common clinical opinion (where SSRIs are viewed as effective antidepressants, and St John's Wort judged as having minimal effectiveness for those with clinical disorders).

Consequences of non-specific findings

Regrettably, non-specific findings encourage some commentators to be highly critical of antidepressant drugs, claiming that such data demonstrate that they either act as placebos or are no more effective than placebos. The situation is not dissimilar for CBT. Despite its high scientific support and cachet value, when we examined (Parker et al., 2003b) the meta-analytic data set for CBT, we returned the Scottish verdict of 'not proven'. In essence, the data set could be interpreted as suggesting that CBT is no more efficacious than any other psychotherapy, counselling, or even placebo psychotherapy; but we suspect that such results more reflect CBT being tested as having universal application and according to limited RCT paradigms (with their emphasis on major depression, brief duration, and on measuring state depression levels). Non-specific findings are not limited to RCTs for 'major depression' however. Himmelhoch (2003) has described how RCT procedures for testing mood stabilisers such as lithium and valproate for bipolar disorder have also counter-intuitively failed to find differences between active drugs and placebo. Himmelhoch concluded (perhaps rather world wearily in relation to large scale, multicentre RCTs) that 'There is no treatment they cannot make equal to placebo.'

8 A 'Declaration of Independence'

We suggest (Parker et al., 2003a) that there are two key reasons limiting the current capacity to interpret whether an antidepressant treatment is efficacious or not, be it a drug, a psychotherapy, or any other modality. Firstly, the use of constructs such as 'major depression' as a standard procedure in such studies. If 'major depression' is, in reality, a heterogeneous non-specific class comprising numerous depressive conditions, each with varying capacities to respond to differing interventions, then any condition–treatment specificity effects will be 'washed out' by homogenising disparate conditions under the 'major depression' rubric. Secondly, by testing treatments as if they have 'universal' (or non-specific) application, rather than as having specific benefits to certain conditions, there is again a diffusion effect. Thus, if a treatment is tested non-specifically against a non-specific condition, why should we not anticipate a meaningless non-specific result?

An analogy: if a woman develops a breast lump and seeks assistance, she is unlikely to be satisfied if the lump is defined in terms of its size (i.e. 'major', 'minor', or 'sub-syndromal'). In essence, she seeks more meaningful information (i.e. is it malignant or benign, does it require treatment, and what are the comparative advantages and disadvantages of differing treatments?). Let us imagine for the moment that breast lump research operated to the same parameters as occurs in the RCTs for testing an antidepressant treatment. Women with 'major lumps' (i.e. size counts, pathology ignored) would be assigned to receive any one of a number of treatments (e.g. radiotherapy, chemotherapy, radical surgery, and lumpectomy). If tested against an 'active' treatment applied non-specifically, differences might or might not emerge, depending on the mix of disorders. For instance, if a high percentage of the lumps were cancerous, we would anticipate that most of the treatments would be efficacious, but differentiation across treatments would be minimal. If most were benign (say, transient cysts), then the active treatments might have marginal efficacy at best, while their appropriateness (in terms of side effects) for those with benign conditions would raise major ethical concerns. Further, if the cysts had a high spontaneous remission rate, the capacity to demonstrate any truly effective treatment would be compromised. If, subsequently, several active treatments were examined across multiple studies using meta-analytic techniques (i.e. viewing each treatment as having universal application and with the mix of conditions and their pathology ignored), would we really anticipate clear differences emerging across the differing treatments? And, if they did, how could we interpret any such differences? Presumably, either that one treatment is universally more powerful than others, or that such differences more reflected

the mix of disorders in the particular studies, and with one treatment being favoured only by a greater representation of conditions showing specific responsiveness to that therapy.

Medicine respects clinical description and diagnosis (particularly in identifying syndromes and classes), as well as the identification of causes and the pursuit of treatment-specificity effects. Such a model allows, for any defined condition, a gradient of effectiveness to be specified for differing treatments. Such a commonsense 'horses for courses' model does not currently underpin the conceptualisation and management of the depressive disorders, and we are all losers as a consequence. Clinicians have disinformation, while patients are further compromised by being at the end of a Chinese Whispers' information line. You may believe that the pharmaceutical industry is the big winner, but we are not so convinced. It must be frustrating (and expensive) for them to have to undertake multiple studies to obtain positive outcomes for licensing reasons, or to have to 'over-power' study numbers to begin to hope for a positive result, or in handling the recent controversies over the efficacy of the SSRI antidepressants. Some might argue that it is in the interests of the industry to have such a situation. This would hold if a pharmaceutical company had a drug (let us call it 'Nirvana') that has universal application. But, and at least up to the time of this book appearing, no such drug exists. The suggestion of the 1990s, that all antidepressant classes have similar effectiveness rates, is no longer believed by clinicians, or by the pharmaceutical industry. Any such claim or imputation has a limited shelf life, as clinicians are disappointed and frustrated by any over-sell. In the same way that most of us would welcome an antibiotic that cured all respiratory disorders, we recognise that certain antibiotics may be very useful for pneumonia, bronchitis, and other specific respiratory conditions but not always useful (and at times counter-productive) for other disorders (e.g. a pulmonary embolus). When antibiotics are used for relevant conditions, their 'usefulness rate' is increased beyond the rate observed when they are used indiscriminately, and their 'specificity' benefit becomes evident. As patients, we thus respect and benefit from the relevant application of those drugs. As clinicians, we appreciate knowing the 'rules of the game' for their prescription and knowing their right ecological niche (i.e. their rational and specific application). Both patients and clinicians, then bless the pharmaceutical company for having an effective product. As in the case of antibiotics, there is thus wisdom in knowing the specific circumstances for prescribing an antidepressant, including the advantages and disadvantages of using a narrow action or a broad action antidepressant drug.

Returning to the limitations of RCTs, there are other reasons why they (at least as the trials are currently undertaken) are delivering meaningless results. In essence, their theoretical advantages have been subverted by procedural components. Recruitment procedures are problematic and increasingly so. Most RCTs (of antidepressant drugs, at least) focus on recruiting relatively pristine subjects, in the sense that any such depressed individuals should not be suicidal, should not have any distinct co-morbid conditions (such as anxiety disorders, drug and alcohol problems, and personality disorders), while a focus on out-patients and volunteers generally excludes those with the more biological disorders such as melancholic depression. Thus, those taking part in studies generally bear little resemblance to depressed patients seen in clinical practice. Further, there is a high natural or spontaneous remission rate in those who take part in antidepressant drug trials, whether they receive the antidepressant drug or placebo. While it used to be in the order of 40%, it is now not uncommon for the responder rate to be 60% in many antidepressant drug trials, with Walsh et al. (2002) quantifying an increase in the responder rate (to both drug and to placebo) of 7% per decade. This suggests that confusion is possibly coming less from a perceived 'placebo response' in such studies and more as a consequence of many participants having either transient or ephemeral depressive conditions, and a high likelihood of a spontaneous remission shortly after baseline assessment. That is, studies are probably being increasingly weighted to those with more 'reactive' disorders who are likely to be primed to respond once the gun has been fired (i.e. received their first set of medication, be it drug or placebo). Thus, we suggest that the RCTs are failing us, in their application. End result, a meaningless database for clinical decision making.

In a recent article, the distinguished psychopharmacologist, Meltzer, observed, after noting several other problematic analyses, that the 'field would appear to have a lot of problems in its scientific basis' but that much of the negative results from studies 'represents the result of not taking all available information into account and a lack of understanding of the importance of specific features of the illness in question' (Meltzer, 2004). Precisely!

Treatment effects

The consequences of a meaningless database are predictable in creating an 'Everyone's a Loser and Everyone's a Winner' conclusion. The lack of differentiation of treatments invites the 'Everyone's a Loser' challenge, that all therapies act non-specifically or via a placebo effect, a rather demeaning

interpretation for those with clinical disorders and a non-inspiring result for psychiatry and for current antidepressant treatments. By contrast, the non-specific database builds to the 'Everyone's a Winner' interpretation, as therapists can readily claim that *their* treatment is efficacious, whatever it actually is. This risks leading to what is described in literary criticism as the 'affective fallacy', where literature (here the therapy) is judged impressionistically rather than by any integral strengths. A hardly surprising consequence if integral strengths resist identification and definition as a consequence of the non-specific model of 'depression'. As noted earlier, this risks a procrustean model where the patient is fitted to the therapist's preferred treatment model – which may reflect the therapist's preferred style, disciplinary background, training or interest – rather than rational logic.

One could argue that this lack of differentiation of specific treatment effectiveness and the 'Everyone's a Winner' interpretation suggest that a single treatment for depression could be disseminated on a large scale at a population-based level. This might be analogous to treating all presentations of shortness of breath in the community with steroid-based inhalers. Whilst many would gain relief, others however (whose symptoms may be related to chronic infection) might actually worsen, while many would receive no benefit as the treatment was inappropriate. The 'average' patient in the community might be better off, but some individuals will be worse off. For the depressive disorders, where disability and suicide are key risks, some individuals will pay a very high price for receipt of a non-specific treatment.

We therefore argue strongly against a dimensional model of depression, the view that antidepressant therapies have universal application, and any model that ignores cause or aetiology. We understand the history, and the important watershed contribution of DSM-III. In the years prior to the development of DSM-III, North American psychiatry had lost its way somewhat, with an emphasis on Freudian psychodynamic treatments moving psychiatry further away from medicine, challenging the credibility of psychiatry and psychiatrists. As noted by Kirk and Kutchins (1992), the publication of DSM-III in 1980 'abruptly shifted emphasis from the aetiological psychodynamic perspective that had dominated psychiatry since World War II'. However, to replace psychodynamic causes with 'other' causes had certain risks. Rather than offend any one group, removing any consideration about aetiology or cause had pre-emptive benefits to the enterprise. In addition, there were 'integrative' models that minimised any logical need to consider cause. In 1973, Akiskal and McKinney had published an article in *Science* that was so influential that a variant was subsequently published in the *Archives of General Psychiatry* (Akiskal & McKinney,