THE PHARMACOTHERAPY AND PHARMACOECONOMICS OF DEPRESSION

STEPHEN TIMOTHY KINGSCOTE-DAVIES

Master of Philosophy

ASTON UNIVERSITY

September 2011

ASTON UNIVERSITY

THE PHARMACOTHERAPY and PHARMACOECONOMICS OF DEPRESSION

Stephen Timothy Kingscote-Davies

Master of Philosophy

2011

Major depression is a potentially life-long mental illness. An episode of depression can last many weeks or months, during which time the patient may suffer disturbance of social relationships, poor work record or inability to work, and possibly be at risk of self-harm, including suicide. Because of the risks associated with depression, the illness needs better recognition by both health professionals and the general public, and a greater understanding that it is an illness that can be treated.

Treatment is primarily by pharmacological intervention, rather through the use of psychological therapies (although for mild depression this is more acceptable and appropriate in many cases). This does mean that such treatment should be at a reasonable cost to the health service with good outcomes for patients and healthcare payer alike, as the treatment is likely to be of long duration, possibly for life. In the cost-constrained health service, value for money is paramount. For the patient, the drug chosen should relieve the symptoms of depression with good tolerability.

The aim of this study is to assess the relationship between the results of randomised controlled trials of antidepressants and the economics of depression and antidepressants. The pathophysiology of depression and the range of antidepressants available in the UK to treat depressive disorder are reviewed. There is currently a debate regarding the use of antidepressants that have so-called "dual action" as to whether these drugs have greater efficacy than those that have only a single (or predominantly single) mode of action.

Randomised controlled trials are the backbone of providing efficacy data, both for licensing approval and subsequently after licensing. However, depression is difficult to measure accurately: there are no 'hard' data such as blood pressure measurements. Rating scales are used to assess the level or depth of depression and its progress during drug therapy. In clinical trials they are essential, while in clinical practice they are perhaps less often used. The common rating scales are discussed, focussing particularly on the two most used in clinical trials (the Hamilton Depression Rating Scale and the Montgomery-Asberg Depression Rating Scale).

Two chapters review and analyse the literature for the pharmacoeconomics of depression and the use of antidepressants, using a recently introduced drug, escitalopram, as an example of the pharmaceutical industry's desire to maintain a drug patent, under the pretext of introducing a valuable new addition to the drug armamentarium for treating this potentially serious, enduring illness.

Keywords: Escitalopram, rating scales, cost effectiveness, meta-analysis.

ACKNOWLEDGEMENTS

I would like to thank my family for their forbearance while this thesis took shape. Also to all my supervisors, Professor Alain Li Wan Po, Professor Femi Oyebode, Professor Martin Griffin and Professor Keith Wilson, for their support and encouragement, to the pharmaceutical companies who have supported my endeavours in various ways, and to the three NHS Trusts who have part supported me. I would like also to extend my thanks to those authors and publishers who have kindly given their permission to use various graphs, figures and diagrams from their published papers.

CONTENTS

LIST of FIGURES AND TABLES

ABBREVIATIONS

5-HIAA	5-Hydroxyindole acetic acid
5-HT	5-Hydroxytryptamine (serotonin)
AMPT	α-methyl-ρ-tyrosine
BDI	Beck Depression Inventory
BDNF	Brain-derived neurotrophic factor
CGI	Clinical Global Impression
CIT	Citalopram
COI	Cost of Illness
CREB	cAMP response-element binding protein
CRF	Corticotrophin releasing hormone
CRS	Carroll Rating Scale
DALY	Disability Adjusted Life Year
DOPA	Dihydroxyphenylalanine
DSM	Diagnostic and Statistical Manual
DST	Dexamethasone suppression test
EPDNS	Edinburgh Post Natal Depression Scale
EQ-5D	Quality of life questionnaire developed by the EuroQoL group
ESC	Escitalopram
fMRI	Functional magnetic resonance imaging
HAMD	Hamilton Depression Rating Scale
HMO	Health maintenance organisation
HPA	Hypothalamic-pituitary axis
ICD	International Classification of Diseases
ICER	Incremental cost-effectiveness ratio
LOCF	Last observation carried forward
LTD	Long-term depression
LTP	Long-term poteniation
MADRS	Montgomery-Åsberg Depression Rating Scale
MAOI	Monoamine oxidase inhibitor
MAR	Missing at random
MARI	Monoamine reuptake inhibitor
MCAR	Missing completely at randon
MMRM	Mixed effects model repeated measures
MNAR	Missing not at random
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
NA	Noradrenaline (= norepinephrine)
NE	Norepinephrine (= noradrenaline)
NICE	National Institute for Health and Clinical Excellence
NNH	Number Needed to Harm
NNT	Number Needed to Treat
NRI	Noradrenaline reuptake inhibitor
OR	Odds Ratio
PET	Positron emission tomography
PPI	Proton pump inhibitors
QALY	Quality Adjusted Life Year
QLDS	Quality of life depression scale
RCT	Randomised controlled trial

RD	Risk Difference
RDC	Research Diagnostic Criteria
RIMA	Reversible inhibitor of monoamine oxidase-A
RR	Rate Ratio
SARI	Serotonin antagonist and reuptake inhibitor
SD	Standard deviation
SE	Standard error
SERT	Serotonin reuptake transporter
SF 36	Short Form 36
SNRI	Serotonin and noradrenergic reuptake inhibitor
SPECT	Single photon emission computed tomography
SSRI	Selective serotonin reuptake inhibitor
TCA	Tricyclic antidepressant
TDO	Trytophan 2,3-dioxygenase

CHAPTER 1

OVERVIEW OF DEPRESSIVE ILLNESS

Introduction

Depression is a potentially chronic, disabling disease, with a high degree of recurrence (Solomon *et al.* 2000). Mental illness, generally, tends to be of a long-standing, chronic nature, an episode of which can last many months or even years before remission of symptoms might occur. When remission or apparent cure do occur, relapse and further periods of illness can be only too common during the individual's lifetime. This pattern of treatment, relapse and remission in depression can be divided into three phases of acute, continuation and maintenance, as shown in Fig 1, which describes the time course of depression (reproduced with permission from DJ Kupfer, 1991). Most episodes of depression last for approximately six months (Üstün *et al.* 2004). Treatment should be continued for at least six months after remission (Anderson *et al.* 2000).

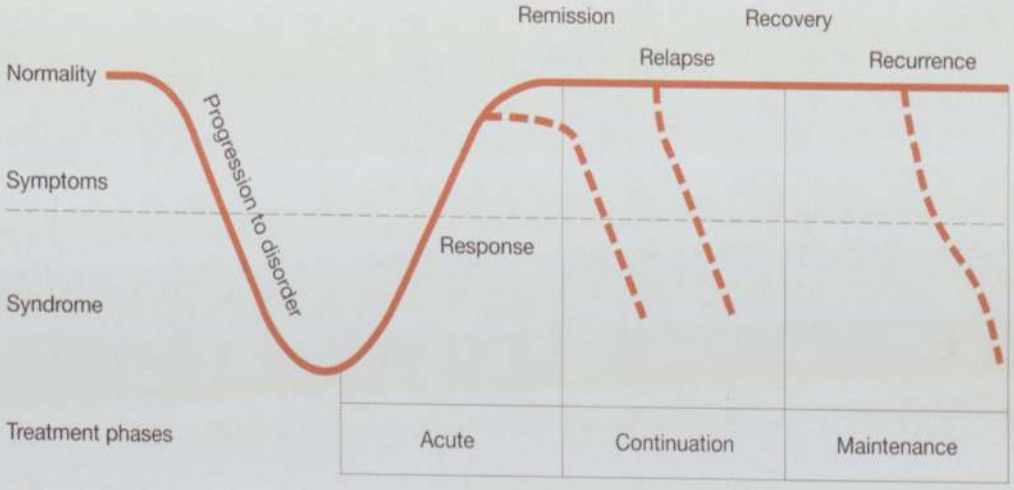


Fig 1: Time course of depression Adapted from Kupfer 1991

Depressed individuals who have had one episode are likely to have a 50% or greater chance of having a second depressive episode, with a probability of further future episodes being 80% to 90% (Kupfer, 1991). During this time there is an increased risk of suicide. Over a potentially considerable period of time, depressive illness can have substantial impact on resource allocation and consumption of those resources, particularly as the quality of life worsens for the individual, when they may make more demands on health care and social services. Patients with more severe depressive disorder will probably require input from many professionals, particularly if hospitalisation is required. There is also evidence that depression can worsen prognosis of medical illnesses, for example, cardiac disease (Jiang, 2001).

It is therefore important to have effective treatments for an illness with such potential long-term consequences but there is a need to establish clear evidence for what those treatments should be, not only for efficacy, but also effectiveness and value for money. The mainstay of treatment is drug therapy, i.e. antidepressants, particularly for moderate to severe depression. Mild illness can be treated with psychotherapy without recourse to drug treatment (NICE, 2009). The antidepressants on the market have proven efficacy in that they are able to treat the majority of patients in randomised controlled trials (RCTs) but can the data from RCTs truly support this for effectiveness, which is the real world situation? For a drug to reach market, it has to undergo rigorous randomised trials, usually against placebo and possibly, but not necessarily, against active controls. But such methods raise ethical issues regarding the use of placebos, the measurement of the depth of depression, the response to placebo, and how to determine the points at which response and then remission are reached.

At the time of embarking on this study, the new antidepressants that were being brought to market had no greater efficacy than their predecessors, although the industry would persuade otherwise, and they had higher acquisition costs. The author felt that there needed to be a way of drawing together the efficacy and ADR data from RCTs, and the economic data that may be available.

Aim of the Study

To determine the relationship between the results of randomised controlled trials in depression and the pharmacoeconomics of depression and the antidepressants.

Objectives

1 Review the current pharmacotherapy of depression.

2 Review the pharmacoeconomics of depression and its pharmacotherapy.

3 Review the introduction of an antidepressant in relation to its efficacy and its economic value.

4 To draw conclusions based on the foregoing analyses.

The patient needs an effective treatment that not only succeeds in relieving their symptoms and their illness but also does so without causing side effects that may lead to the patient ceasing to take medication. If this cannot be done, it prolongs the illness, the disease burden for patient, family and ultimately society, and raises the costs of treating the illness. Chapter 2 will briefly examine the aetiology of depression and its treatment with

antidepressants to put the research into context of a complex illness with no real advances in its treatment.

It is essential to develop tools to estimate the depth of illness, both in the clinical setting and in clinical trials. A number of rating scales have been developed over the last forty years to assess the severity of depressive illness, but it is the oldest that are still used in clinical trials. These last issues will be investigated in Chapter 3 that reviews the rating scales used in clinical trials.

The drugs available for treating depression until the late 1980s had low acquisition costs and were usually quite effective. But many patients would only gain partial relief from symptoms, and suffer relapse or recurrence of the illness due to a lack of efficacy, or poor compliance. The older drugs have a wide range of pharmacological actions giving rise to various adverse effects, which can lead to non-compliance with the therapy due to poor tolerability. This may lead to lower than effective dosing, which might account for the lack of efficacy (and therefore relapse) or only partial efficacy. Side effects would not necessarily be less with lower doses, although this might be the intention of the prescriber. Depression is a costly disease, responsible for a large burden on society and the nation, as well for individuals. If treatment is not effective in resolving the illness, either through lack of effectiveness or intolerability, then that treatment does not give value for money. The pharmacoeconomics of depression and its treatment have been investigated before, but Chapter 4 attempts to distil the available data and discusses the relative merits of some key economic analyses.

The recent introduction of escitalopram is used to illustrate how a drug can be brought to market with allegedly poor data. When escitalopram was launched in 2002 by Lundbeck/Forest, much was made of its efficacy, based on the pharmacology of the molecule. The supporting data, particularly one pooled analysis, was considered by many to be poor. Chapter 5 briefly reviews the pharmacology of escitalopram and discusses the rationale behind bringing the drug to market. All appropriate randomised controlled trials are included in a meta-analysis of its efficacy, as at its launch and subsequently, there was no comprehensive such analysis.

Chapter 6 will bring together the available data as discussed in the previous chapters and attempt to formulate a way forward for the ideal trial in mental health.

CHAPTER 2

AETIOLOGY OF DEPRESSION
and
PHARMACOLOGY of ANTIDEPRSSANTS

Introduction

This chapter will review the pathophysiology and the drug treatment of depressive illness. An understanding of depression and the probable multifactorial reasons for an individual becoming depressed underpins the background of this study.

Depression: the illness

History

For over 2000 years, man has been trying to understand how the human body functions, not least the brain. Hippocrates (460-357 BC) considered the possibility that external factors, such as planetary conjunctions, would cause the spleen to excrete black bile which would lead to alterations in mood. There were no significant discoveries between then and the 20th century as medicine had no sophisticated biochemical or biophysical techniques to call upon. However, Robert Burton published a book in 1621 entitled 'Anatomy of Melancholy' which discussed the possibilities of heredity, the influence of alcohol, diet and biological rhythms. Over 200 years later Emil Kraepelin (1856-1926) formulated the view that there was a genetic contribution to manic-depressive illness: he also hypothesized that there are morphological changes in the brain, although post-mortem studies did not prove anything.

> *That ...inbred cause of Melancholy is our Temperament, in whole or part, which we receive from our parents...it being a hereditary disease; such as the temperament of the father is, such is the son's, and look what disease the father had when he begot him, his son will have after him...And that which is more to be wondered at, it skips in some families the father, and goes to the son, or takes every other, and sometimes every third in a lineal descent, and doth not always produce the same, but some like, and a symbolizing disease..."*
> Robert Burton
> *The Anatomy of Melancholy* (1621)

Others have taken a more psychological view based on work by psychoanalysts such as Freud and Jung. However, Adolf Meyer (1866-1950) further developed these theories into a psychobiological theory in which he postulated that environmental factors might play a part. That is an individual might be born with a genetic predisposition to depression but that the illness would only manifest if there were external factors that acted as triggers to its development.

Although psychoanalytic theory and practice were prominent post World War II, during the 1950s biological theories of psychiatric illness became more established, especially after the introduction of chlorpromazine and then the antidepressants. It was during the 1960s that researchers such as Schildkraut, and Bunney and Davis developed the monoamine hypothesis which basically states that a relative lack of monoamine would lead to depression, a relative excess to mania. Over the last few decades, sophisticated imaging techniques, animal model paradigms, the discovery of other potential neurotransmitters and the role of stress and glucocorticoids have improved our understanding.

Pathophysiology

In 1937, Papez proposed the limbic system to be the 'seat of human emotions' (cited by Musselman in Textbook of Psychopharmacology), but for many years research was restricted to examination of post-mortem brains and animal studies using paradigms of depression. The introduction of computed tomography (CT) and subsequently more sophisticated neuroimaging techniques, such as magnetic resonance imaging (MRI), has led to imaging of human brains in vivo. Functional studies using positron-emission tomography, single photon emission computed tomography and functional MRI show not only the anatomy and structure of the brain but also the function of different regions.

The hippocampus is a vital component of learning and memory, control of emotion, regulation of the hypothalamic-pituitary-adrenal axis, and other vegetative processes. It is one of only a few regions of the brain which has dynamic neuronal growth and plasticity throughout life (Malberg *et al.* 2000). The various cognitive and somatic symptoms that may be seen in depression might be, at least in part, explained by structural or functional changes in the hippocampus. Animal studies have shown that prolonged exposure to large doses of corticosteroids can lead to permanent loss of hippocampal neurons (Duman *et al.* 1999). The usual feedback loop of the hypothalamic-pituitary-adrenal axis and the hippocampus appears to break down, resulting in raised glucocorticoid levels and further damage to the hippocampus (Bremner 2002). Hippocampal atrophy has been demonstrated in patients with recurrent major depression (Sheline *et al.* 1996; 2003).

In addition, it has been noted that the longer the state of depression in patients, the greater the extent of hippocampal atrophy (Campbell & Macqueen 2004; Sheline *et al.* 2003; Neumeister *et al.* 2005) (see Fig 2).

The MRI studies performed so far have generally not included information as to the degree or depth of depression of the subjects. This would be useful to include in future studies, as it would help to explore the correlation between the changes in the hippocampus and the response to antidepressant treatment. For example, three studies reported in Videbech & Ravnkilde (2004) showed either smaller volume in the right hippocampus (two studies) or reduced density in the left (one study) and these changes were linked to poor response to antidepressant medication. If confirmed, these results would have implications for predicting a clinical response. However, the studies performed so far have been cross-sectional and they cannot answer the question of the decreased hippocampal volume being caused by depression or whether the reduction in volume predicts the

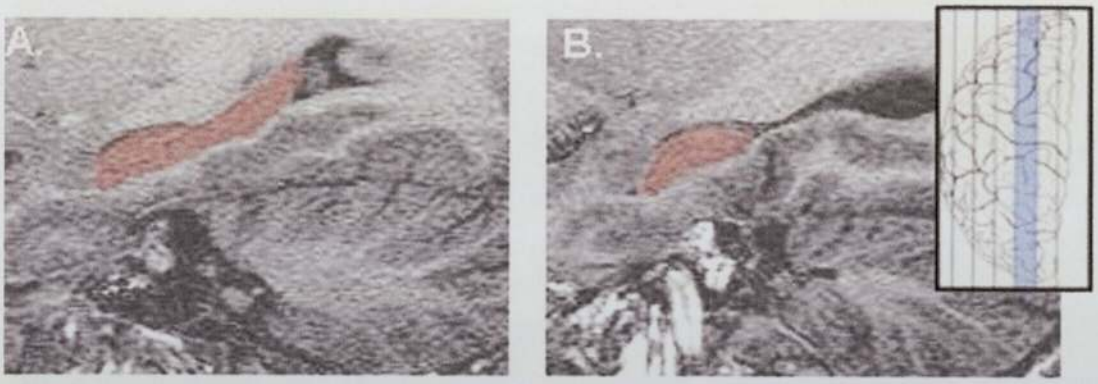


Fig. 2: Magnetic resonance spectroscopic images of the left hippocampus in a healthy control subject and in a patient with recurrent depression. The size of the difference shown here is unusually large, with most positive studies reporting a reduction in hippocampal complex (HC) volume of about 15% between cases and controls. Insert shows in blue the approximate sagittal level of the HC. Images were acquired on a 1.5-T GE Sigma Genesis–based EchoSpeed imager using previously published parameters. A: Sagittal view of the left HC, highlighted in red, of a healthy control subject whose left HC volume measured 3295 mm^3. B: The patient whose left HC is represented here, with an HC volume of 2015 mm^3, was of the same age and sex as the control subject but had a long history of recurrent depression. (Images courtesy of Dr Glenda MacQueen, McMaster University)

development of illness. Not only are longitudinal studies required but they should also combine the measurement of hypothalamic-pituitary axis (HPA) activity with measurement of hippocampal volume. This would help elucidate the relationship between the dysregulation of the HPA axis and the amount of hippocampal loss.

Given a certain level of stressful life events, individuals will differ in their response. A certain proportion of the population will develop depression. These individuals may have a genetic predisposition to development of a vicious cycle of cellular events in which increased cortisol levels gradually overstimulate the cells of the hippocampus, leading to cell death. This has the consequence that the inhibitory regulation from the hippocampus on the HPA axis is further decreased and raises cortisol and corticotrophin releasing factor (CRF) levels. This appears to be reversed by antidepressants, demonstrated

by the administration of imipramine to rats, which upregulated glucocorticoid receptors in the hypothalamus and hippocampus and reduced the overall activity in the HPA (Kitayama *et al.* 1997).

It is not only the hippocampus that is affected by these changes but other regions of the brain are as well, including the prefrontal cortex, amygdala, striatum and thalamus. Depending on how neuronal connectivity and activity are affected in these regions, and to what extent, these changes will probably determine the expression of depressive disorder and may, perhaps, ultimately account for the subtypes of depression seen clinically. The hypothalamic pituitary adrenal axis is involved in stress response and the consequent influence of glucocorticoids on brain regions that are integral for maintaining mood is important. One to two thirds of depressed patients show signs of a hyperactive HPA axis, with either elevated glucocorticoid levels (due to hypersecretion of corticotrophin releasing hormone) or a positive dexamethasone suppression test (DST); ie dexamethasone does not suppress cortisol levels as it would in normal subjects (Dinan 1998). The DST is sensitive for depression but it is not specific for it as non-depressed subjects with alcohol dependence, anorexia nervosa and early Alzheimer's disease also show abnormalities in cortisol secretion.

There are two broad types of pathway involved with signal transduction within the cell. The first includes pathways that are usually regulated by the 'classic' neurotransmitters (monoamines) through receptor-coupled second messengers such as cAMP. The second pathway includes intracellular systems controlled by receptors containing or interacting with protein tyrosine kinases, regulation of which is usually by neurotrophic factors and cytokines. These pathways are vital to the control of all aspects of neuronal function and underpin the adaptability and response of the brain to various chemical and environmental inputs. Such changes may result in changes of synaptic activity or morphological changes in brain structure that may be beneficial (eg sprouting of neurons), or adverse, including atrophy.

Substance P

Substance P (SP) was discovered over 70 years ago but at the time its function was unknown (DeVane 2001). Since then, SP has become the best understood and most intensively studied of the neuropeptides. In the 1950s, SP was considered to be the pain transmitter for primary afferent sensory fibres, probably concentrated in the dorsal roots of the spinal cord. After its purification and subsequent synthesis in the 1970s, further

research demonstrated the role of SP as a potentiator of excitatory inputs to nociceptive neurons.

To exert its effects, SP binds most strongly to the NK-1 receptor (the SP receptor) The SP receptor is a G-protein-coupled receptor that then activates several second messenger systems. After release from nerve terminals, SP is rapidly degraded by several proteases.

In the CNS, SP-containing neurons are found in distinct brain regions such as midbrain and basal ganglia, hypothalamus, the limbic system including the hippocampus and amygdala, and the spinal cord. SP is co-localised with other neurotransmitters and modulates the effect of neurotransmission (Herpfer 2005). These regions are important to the regulation of mood and the neurochemical responses to stress so it appears that there may be a link between SP and affective disorders (Bondy *et al.* 2003).

Brain-derived neurotrophic factor

Brain-derived neurotrophic factor (BDNF) promotes the function, sprouting and regrowth of serotonergic neurons in adult rat brain. Using depression paradigms has shown that infusions of BDNF into the dorsal raphe nucleus have an antidepressant effect, while various forms of stress in rat depression paradigms decrease the amount of BDNF expression in the rat (Smith *et al.* 1995). Atrophy and loss of hippocampal neurons may be attributable to this decrease in the amount of BDNF, but this model may only refer to some subtypes of depression (Duman *et al.* 1997). It has been shown that chronic administration of antidepressants reverses this effect by increasing the levels of a mRNA coding for a transcription factor, (cAMP response-element binding protein, CREB), and the levels of the protein itself (Duman *et al.* 1997). Electroconvulsive stimulation in the laboratory (clinically used as ECT) has been shown to increase CREB mRNA levels (Nibuya *et al.* 1996), supporting the concept of CREB being an important part of maintaining cell function and integrity.

It takes 14-21 days to induce CREB production; a similar time for the expression of antidepressant effect (Nestler *et al.* 1989) and there is a similar delay in the expression of BDNF. The administration of antidepressants may indirectly increase the synthesis of BDNF and thereby improve repair of neurons. Also, the neurodegenerative process in the hippocampus can be reversed by electroconvulsive stimulation, which promotes the expression of BDNF and thus increases growth of neurons in the hippocampus (Vaidya 1999).

Could BDNF dysregulation be an underlying aetiology for depression? If so, it might (in part at least) explain the common two-week delay of antidepressants. If they are indirectly increasing neuronal regrowth, this is a slow process.

The Depression Gene

Family studies in the USA have demonstrated an hereditary component to depression, relatives of depressed patients being at higher risk than the general population, (Johansson *et al.* 2001). The relative risk of first-degree relatives having depression (relative to the general population) is 2 to 3 (Levinson 2005). Depression probably arises due to a combination of genetic and environmental factors but very little is yet known about which genes are involved. There may be a genetic predisposition in some individuals that, under the right circumstances, leads to depression. Conversely, others are relatively resistant to developing the illness (Duman *et al.* 1997).

It appears that critical genes may be down-regulated by stress, possibly via monoamine neurotransmission, resulting in a reduction of the gene products. One particular part of the signal transduction process which is being studied is the target gene for BDNF, but it is important to note that there is most likely to be more than one gene responsible for the failing in signal transduction: i.e. that multiple genes on the genome are involved. The problem is to determine which ones are important and how they interact with each other to either protect against depression or lead to its emergence. That there are several stages in signal transduction and cellular integrity (cAMP, CREB, BDNF) leads to the conclusion that an enormous number of genes are probably involved (Stahl 2000a). Critical genes that may be involved with coding for neurotransmitter receptors may be activated or deactivated by antidepressants (Stahl 2002b).

Environmental factors are important as they may predispose an individual to depression later in life. A small amount of neuronal damage may result from exposure to stress earlier but this is not enough to precipitate a full illness. Further insults to the brain may act as triggers to develop into depression.

Monoamine Hypothesis

However, the most widely accepted theory, the monoamine hypothesis, suggests that depression arises from reduced concentrations of norepinephrine, serotonin and/or dopamine in the synapse. It was observed in the early 1950s that the antihypertensive drug, reserpine, caused depression in some patients. It was found that reserpine depleted brain

serotonin stores and increased concentrations of the serotonin metabolite, 5-hydroxyindoleacetic acid (5-HIAA) in urine (Schildkraut 1969; Maas 1975). With presynaptic stores of NA and 5HT depleted, there is less for release into the synapse. The depressive symptoms produced were reversible on stopping the reserpine. In addition, administration of the NA precursor, dihydroxyphenylalanine (DOPA), was found to be effective at reversing reserpine-induced changes in an animal model of depression. This finding was also replicated in humans.

Further evidence came from the observation that the TB drug, iproniazid, improved mood in TB patients who also had depression. In 1952, iproniazid was demonstrated to inhibit monoamine oxidase (MAO), the mitochondrial enzyme responsible for metabolising monoamines in the presynaptic terminal. This increases the availability of NA and 5HT in the presynaptic terminal for release across the synapse. Later it was noted that non-TB depressed patients could also be treated with iproniazid. During the 1950s, other similar compounds were developed; in particular, isocarboxizid, phenelzine, tranylcypromine. Unfortunately, this hypothesis was proposed in reverse: i.e. the clinical findings developed a hypothesis for a disease process instead of original biochemical and neurophysiological research establishing a defect in a physiological mechanism.

However, this hypothesis has been supported by studies in the 1990s that examined the effects of reducing levels of NA and 5HT to elucidate the roles of these transmitters in depression (Miller *et al.* 1996a; 1996b). One study in depressed patients demonstrated the therapeutic effects of a reuptake inhibitor could be reversed by depletion of the neurotransmitter it affected. In the study, patients were assigned to either desipramine or fluoxetine treatment arms. After remission was achieved, responders were given α-methyl-p-tyrosine (AMPT), which blocks NE synthesis. Most (81%) desipramine responders relapsed. This contrasted with the fluoxetine responders, only 19% of whom relapsed. Other data looking at serotonin depletion induced by amino acid drinks to deplete tryptophan showed a higher relapse rate in the SSRI remitted patients than those who had been given desipramine. Other work has also demonstrated this (Charney 1998). Reductions in brain serotonin and its major metabolite 5-hydroxyindole acetic acid (5-HIAA) have been noted in the brains of post-mortem depressed and suicide patients (Owens & Nemeroff 1994). Reduced concentrations of 5-HIAA in the cerebrospinal fluid of drug-free depressed patients have been found, and if patients in remission who have responded to a serotonergic antidepressant are given either diets low in tryptophan or tryptophan-free amino acid drinks so that brain concentrations are depleted, they suffer a profound relapse. Thus the depletion of noradrenaline or serotonin appears to lead to

depressive symptoms. The consequence of neuronal monoamine depletion is an increase in the number of post-synaptic receptors, *up-regulation,* the temporal relationship of which correlates well with the onset and development of depressive illness. Giving an antidepressant over a period of time causes the receptor numbers to decrease, *down-regulation.* The temporal relationship of this correlates well with the well-known delay in the clinical situation for onset of relief of depression of around two weeks with most antidepressant compounds, in spite of the rapid inhibition (hours, *in vivo*) of the reuptake transport systems, so a pure pharmacological explanation as proposed by the monoamine hypothesis is not enough to explain what manifests clinically as depression. If the immediate pharmacological action of antidepressants is not the direct cause of relieving depression, then it infers that there are secondary neurophysiological effects resulting from the administration of antidepressants that account for the actual beneficial clinical effects. There is no direct evidence for primary abnormalities in the brain monoamine pathways (Nestler 1998), however. Taking these data together gives the suggestion that the action of monoamines is the first part of a process which modulates other neurological systems that are more directly involved in the development of depression (Heninger *et al.* 1996).

Early studies showed that the density of $5HT_2$ receptors and β-adrenergic receptors for serotonin and norepinephrine respectively was reduced on long-term antidepressant treatment in limbic brain regions such as the hippocampus and cerebral cortex (Duman *et al.* 1997). One hypothesis based on this suggested that depression might arise from super-sensitivity of these receptors. Down-regulation would therefore be anti-depressant. However, down-regulation occurs at a much faster rate than the speed of onset of antidepressant effect. Another objection to this theory is that electroconvulsive therapy up-regulates $5HT_2$ receptors. However, continued treatment, although down-regulating receptors, does increase the availability of monoamine in the synapse. It is possible that in spite of fewer receptors post-synaptically, neurotransmission is still effective and therefore increases cAMP levels above those found in the no treatment state (Duman *et al.* 1997).

Another mechanism may also account for the slow onset of antidepressant drug action. Since the synapse is not a closed system, when 5-HT is released from the presynaptic nerve terminal, some neurotransmitter escapes the synapse and activates somatodendritic 5-HT_{1A} and presynaptic terminal 5-HT_{1B} receptors. These are part of a feedback mechanism to reduce further release of serotonin from the nerve terminal. When drug therapy is initiated, there is an increase in serotonin levels that activate presynaptic $5HT_{1A}$ autoreceptors. These dampen the rate of neuronal firing from the cell body and also serotonin release from the presynaptic terminals (Salter 1996). Continued treatment down-

regulates these autoreceptors, and reduces the inhibition of neuronal firing and serotonin release. So acute treatment with selective serotonin reuptake inhibitors cannot raise forebrain 5-HT levels to a concentration sufficient for an antidepressant response.

This theory of delayed response has been challenged (Posternack & Zimmerman 2005). The early findings of Khun and others in the late 1950s that imipramine tended to give a therapeutic response within a few days still hold true, in spite of animal experiments and other research describing a delayed response. Posternack and Zimmerman postulate that if there is a delay in response, there would be little or no separation between placebo and active, and that subjects who respond to active medication will do so later than those on placebo. In their review of 47 trials, they found that about 50% of patients respond during the first two weeks of a trial. Such findings have been demonstrated with newer antidepressants such as venlafaxine and mirtazapine. This does not explain experimental data showing desensitisation of inhibitory $5HT_{1A}$ receptors on raphe neurons and changes in sensitivity of α_2 autoreceptors, which may account for the lag-time before symptoms improve (Nutt 2002). To be able to demonstrate clearly an early onset of clinical effect, separation of the investigative drug from placebo and active comparator is necessary during week one, which is then sustained while the comparator 'catches up'.

Taking the monoamine hypothesis a step further, monoamine receptors may have an important role. Whether as a consequence of lower than normal amounts of neurotransmitter, of abnormalities in the post-synaptic receptors, or of abnormalities of the transduction from receptor to processes downstream from the receptors, there is a malfunction of post-synaptic receptors. There appears to be an increased number of them (up-regulation) to apparently compensate for deficiencies in the transmission process.

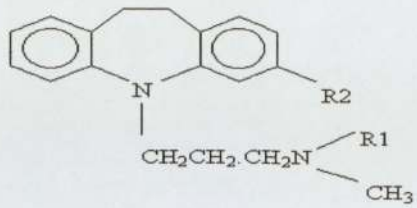
Other work has shown that dysregulation of neurotransmitters, neuropeptides and prostaglandins, or changes in receptor sensitivity may contribute to the pathophysiology of depression (Leonard 1996; Heninger *et al.* 1996; Stahl 2000c). Resistant depression is not particularly amenable to modulators of monoamine activity and combination drug therapy may have to be used, including hormonal treatment (liothyronine), to obtain a response (Aronson *et al.* 1996; Goodwin *et al.* 1982). That some forms of depression are difficult to treat with enhancers of monoamine activity and the fact that electroconvulsive therapy is highly effective in such cases suggest that mechanisms are involved other than that suggested by the monoamine hypothesis. Also patients differ in their response to drug treatment that may enhance either noradrenergic or serotonergic function, in spite of appearing to have the same symptoms. This may indicate different sub-groups of disease (Charney 1998).

The serotonin$_{1A}$ receptor sensitivity hypothesis suggests that there is increased functioning of post-synaptic 5HT$_{1A}$ receptors in the hippocampus, either by increased sensitivity of those receptors or that 5HT autoreceptors are desensitised. So far there have been no specific full agonists of these receptors to test the hypothesis, only the partial agonists buspirone and gepirone. It may also be that more than one receptor type is needed to activate the post-synaptic signalling systems, and there is currently debate as to whether one or two mechanisms are necessary for a full response.

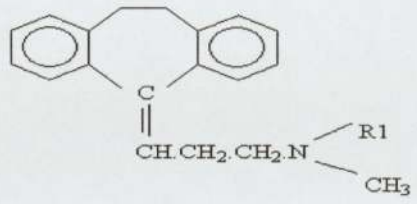
Pharmacotherapy

Figure 3 shows the structures of some of the antidepressants that have been in use over the past four decades. They can be classified into eight groups (Table 1), a classification scheme which is primarily based on their pharmacological action, except for the tricyclic antidepressants (TCAs). The TCAs are named after their chemical structure rather than their pharmacology. To follow the current nomenclature of pharmacological action, the TCAs might be better named as MARIs; monoamine reuptake inhibitors. All current antidepressants modulate the activity of monoamine neurotransmitters, principally NA and serotonin (Nutt 2002) and, notwithstanding some variation, are essentially 'me-too' drugs. Bupropion, not available as an antidepressant in the UK, inhibits the reuptake of dopamine as well as norepinephrine although the inhibition for both monoamines is quite weak (Horst & Preskorn 1998; Learned-Coughlin 2003). It is thought to act more like a pro-drug as it is metabolised to a more potent reuptake inhibitor that is concentrated in the brain (Stahl 2000c).

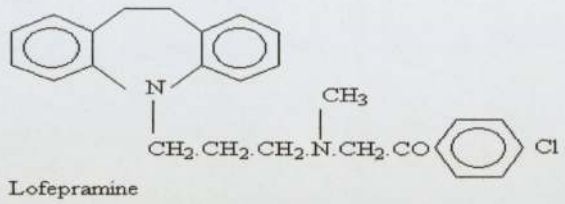
Given the discovery that norepinephrine (NE) levels were decreased in depression, the original emphasis was therefore on increasing the concentration of NE in the synapse, but work with clomipramine in particular showed that serotonin was also an important neurotransmitter for regulating mood, and therefore the serotonin reuptake mechanism would be another target for drug therapy. The drugs have varying degrees of reuptake inhibition potency, as shown in Table 2, which also shows the differing binding affinity profiles for other receptor sites.



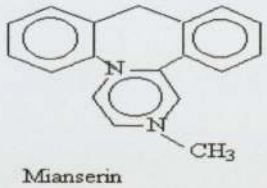
R1= CH_3; R2= H : imipramine
R1= H; R2= H : desipramine
R1= CH_3; R2= Cl: clomipramine



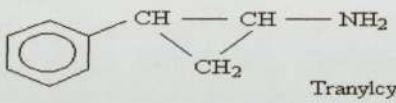
R1= CH_3 : amitriptyline
R1= H : nortriptyline



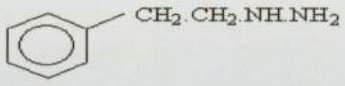
Lofepramine



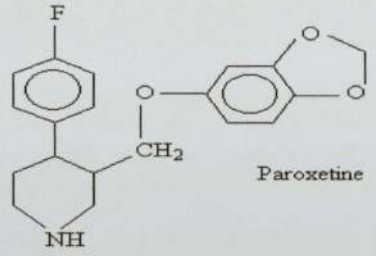
Mianserin



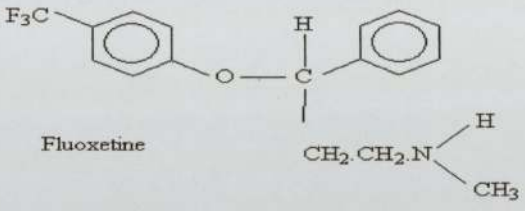
Tranylcypromine



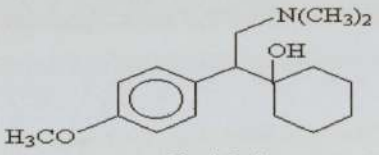
Phenelzine



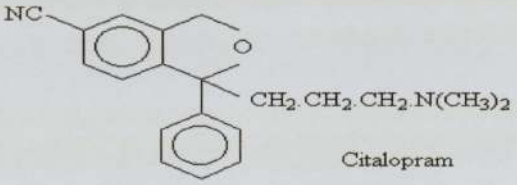
Paroxetine



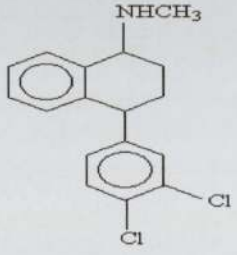
Fluoxetine



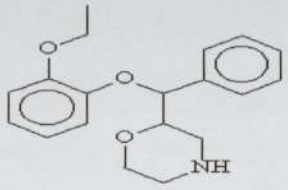
Venlafaxine



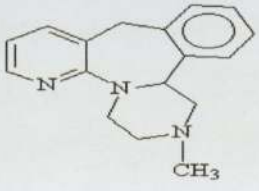
Citalopram



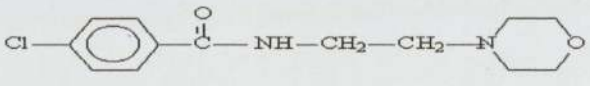
Sertraline



Reboxetine



Mirtazapine



Moclobemide

Fig 3 Chemical Structures of Antidepressants

If the effects on monoamine reuptake are calculated as selectivity ratios for serotonin and noradrenaline, it can be seen there is a wide spectrum of selectivity (Fig 4). But the monoamine hypothesis does not describe the whole story of the relief of depression by antidepressants. In common with the rapid action of many other drugs (antihypertensives, for example), there is an almost instant pharmacological effect of reuptake inhibition, enzyme inhibition or receptor blockade, but this does not correlate with the lifting of depression. Contrast this with the antagonist action at histaminic and

Table 1 Classes of Antidepressants

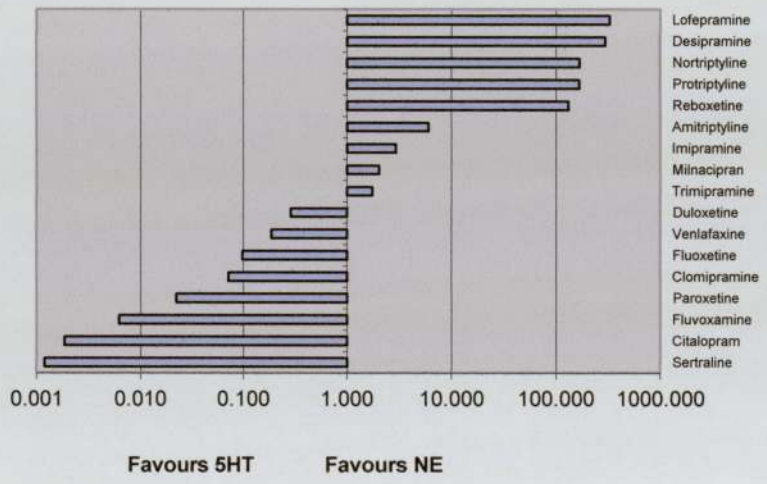
DRUG CLASS	Pharmacology	Examples	Adverse Reactions
TCA Tricyclic Antidepressant	Monoamine reuptake inhibition, primarily NA, but also 5HT	Amitriptyline, Imipramine, Lofepramine, Trimipramine, Dosulepin	Anticholinergic, antihistaminic, cardiovascular
MAOI Monoamine oxidase inhibitor	Monoamine oxidase inhibition	Phenelzine, Tranylcypromine	Sweating, dry mouth, postural hypotension, 'cheese' reaction.
RIMA Reversible Inhibitor MAO	Reversible inhibition of monoamine oxidase-A	Moclobemide	Insomnia, headache, dizziness, nausea
SSRI Selective Serotonin Reuptake Inhibitor	Selective inhibition of serotonin reuptake transporter	Citalopram, escitalopram, Fluoxetine, Fluvoxamine, Paroxetine, Sertraline	Headache, nausea, diarrhoea, somnolence, insomnia, sexual dysfunction, tremor rashes
NaSSA Noradrenaline & Selective Serotonin Antidepressant	Monoamine oxidase inhibition blockade and $5HT_2$ and $5HT_3$ receptor blockade	Mirtazapine	Drowsiness, increased appetite, weight gain
NARI Noradrenaline Reuptake Inhibitor	Selective NA reuptake inhibition	Reboxetine	Dry mouth, constipation
SARI Serotonin Antagonist & Reuptake Inhibitor	Selective serotonin reuptake inhibition and $5HT_2$ blockade	Trazodone (nefazodone)	Headache, dizziness, somnolence, nausea
SNRI Serotonin & Noradrenergic Reuptake Inhibitor	NA and serotonin reuptake inhibition	Venlafaxine, Duloxetine, Milnacipran	Hypertension, nausea, headache, dizziness, somnolence, insomnia, sexual dysfunction

muscarinic receptors. Side effects due to blockade of these receptors can appear rapidly. There is up-regulation of post-synaptic receptors as a response to stress but these receptors are possibly not fully coupled to the intracellular transduction pathways that are responsible for cellular repair.

Antidepressants ultimately down-regulate these receptors. The increased concentrations of serotonin and noradrenaline at the upregulated receptors may account for the initial side effects attributable to these monoamines.

These increased concentrations of monoamines also operate a negative feedback loop via presynaptic α_2 receptors or $5HT_{1A}$ receptors to reduce the firing rate and release of the respective neurotransmitter. However, over a period of four weeks these auto-receptors desensitise and cell firing slowly recovers. This slow recovery raises monoamine levels, which then remain elevated (Nutt 2002), and lends credence to the slow onset of action. These α_2 autoreceptors are targets for mianserin and mirtazapine.

Fig 4: Relative selectivity of 5HT:NE reuptake



Drug	Binding Affinity					Reuptake affinity/potency			
	H1	Musc	Alpha 1	5HT2	D2	5HT	NA	DA	Reference
TCAs									
Amitriptyline	91	5.6	3.7	nd	0.1	1.5	4.2	0.043	Richelson 96
	4	3.5	18	15		nd	nd	nd	Taylor 95
Clomipramine	3.2	2.7	2.6	nd	0.53	18	3.6	0.056	Richelson 96
	54	67	60	54	430	1.5	21	4300	Hyttel 94
Imipramine	9.1	1.1	1.1	nd	0.05	2.4	7.7	0.02	Richelson 96
	25	36	39	68		76	51	nd	Taylor 95
Lofepramine	nd	nd	nd	nd	nd	880	2.7	3300	Hyttel 94
Nortriptyline	10	0.67	1.7	nd	0.083	0.38	25	0.059	Richelson 96
Trimipramine	370	1.7	4.2	nd	0.56	0.04	0.2	0.029	Richelson 96
SSRIs									
Citalopram	470	2900	4500	>1000	>10000	2.6	3900	nd	Brunello 98
	350	5600	1600	5600	33000	108	6100	40000	Hyttel 94
Fluoxetine	0.016	0.05	0.017	nd	0.015	8.3	0.36	0.062	Richelson 96
	nd	nd	>100,000	4210		5	180	nd	Taylor 95
	3200	3100	14000	710	32000	6.8	370	5000	Hyttel 94
Fluvoxamine	0.00092	0.0042	0.013	nd	0.13	14	0.2	0.02	Richelson 96
Paroxetine	0.0045	0.93	0.029	nd	0.0031	136	3	0.059	Richelson 96
Sertraline	0.0041	0.16	0.27	nd	0.0093	29	0.45	0.39	Richelson 96
OTHERS									
Mirtazapine	9.3	6.2	6.5	8.2	<6	-	-	-	Kasper 97
Nefazodone	0.0044	0.0091	2.1	nd	0.11	0.73	0.18	0.042	Richelson 96
	800	14,800	110	22		149	160	nd	Taylor 95
Reboxetine	>1000	>10000	>1000	>1000	>1000	1070	8.2	nd	Brunello 98
Trazodone	0.29	0.00031	2.3	nd	0.26	0.53	0.02	0.007	Richelson 96
	640	66,500	17	7		92	>1000	nd	Taylor 95
Venlafaxine	0	0	0	nd	0	2.6	0.48	0.019	Richelson 96
Duloxetine	2300	3000	8300	504	14000	4.6	16	369	Bymaster 01

Table 2. Receptor Binding Affinities and Potencies at Reuptake Binding Sites

Richelson values: affinities as 10-7x1/Kd, where Kd is the equilibrium dissociation constant in molarity; potency as 10-7x1/Ki, where Ki is the inhibitor constant in molarity (the larger the value, the more potent the effect). Hyttel, Brunello & Moret values are Ki in nM (the smaller the value, the more potent effect). Taylor, Bymaster values are IC50 values in nM (the smaller the value, the more potent). Kasper values are pKi, the negative log of drug concentration causing 50% occupancy (the larger the value, the more potent). nd=not done

Tricyclic Antidepressants (TCAs)

The main antidepressant group for many years has been the tricyclic antidepressants (TCAs). This nomenclature was based on the chemical structures, having a 3-ring basic structure with differing side-chains. These can make a large difference in reuptake specificity: tertiary amines like impramine are more potent inhibitors of serotonin reuptake, while secondary amines like desipramine are more potent inhibitors of norepinephrine reuptake.

The first TCA synthesised was the dibenzazepine compound imipramine in 1948 and was originally developed as a potential antihistamine. When this was not realised, it was tested for its potential as an antispychotic before its ultimate recognition as an antidepressant (Nutt 2002). Following this, amitriptyline, desipramine (the main metabolite of imipramine and a selective NE reuptake inhibitor), nortriptyline, protriptyline and doxepin were developed. These compounds are all monoamine reuptake inhibitors (MARIs) which inhibit either NA or serotonin reuptake or both in the pre-synaptic terminal. The degree to which monoamine reuptake is blocked varies: desipramine and lofepramine are potent inhibitors of NA reuptake, while clomipramine preferentially blocks serotonin reuptake (Table 2, Fig 4). These drugs have actions at other receptor sites which can lead to unwanted adverse effects (Table 1).

All have the same degree of efficacy, and an onset of action of approximately two weeks. Later MARIs, dothiepin (dosulepin) and lofepramine, were no different in these respects but lofepramine has a better toxicity profile (Lancaster & Gonzalez 1989; Pugh *et al.* 1982). All the older MARIs are cardiotoxic in overdose, particularly dosulepin, due to quinidine-like membrane stabilisation. This can be serious for patients with pre-existing cardiac problems as it may precipitate bundle branch or complete heart block. Lofepramine is generally regarded as being less cardiotoxic (Belz *et al.* 1983; Gokelma, 1983). However, lofepramine can cause changes in liver function, which may lead to frank liver damage and toxicity (Committee on Safety of Medicines, 1988).

Many of these side-effects can be avoided or minimised by dose titration, but the problem with titration is that general practitioners may never reach a therapeutic dose (Beaumont *et al.* 1996), while patients may default with medication due to lack of initial effect.

Monoamine Oxidase Inhibitors (MAOIs)

In 1951, isoniazid and iproniazid were developed for the treatment of tuberculosis. It was soon found that iproniazid in particular had mood-elevating effects in patients and

this led to its investigation as an antidepressant in 1952 (Nutt 2002). In the same year, it was found that iproniazid was an inhibitor of monoamine oxidase (MAO). In 1957, it started to be used in psychiatry for depression. Three drugs of this class remain available in the UK, phenelzine, isocarboxazid and tranylcypromine, although the Joint Formulary Committee for the British National Formulary now advises they are less suitable for prescribing (BNF 2005).

The MAOIs irreversibly inhibit both isoforms (MAO-A and MAO-B) of the enzyme responsible for metabolising NA, dopamine and serotonin. Monoamine oxidase is located in the outer membrane of presynaptic neuron mitochondria, where it is the principal enzyme for catabolising monoamines. By preventing the breakdown of monoamines in the presynaptic terminal, more monoamine is available for release into the synapse.

As MAO is important in the periphery for metabolising naturally occurring exogenous sympathomimetics, blocking this enzyme can lead to potentially serious adverse effects. This is the cause of the 'cheese reaction', a hypertensive episode characterised by symptoms such as headache, dizziness, facial flushing, and tachycardia. This can occur after ingestion of foods containing tyramine, whose pressor effects arise from its direct and indirect sympathomimetic actions. Sympathomimetic drugs, such as phenylpropanolamine and pseudoephedrine, found in some cold remedies are also metabolised by MAO. Inhibition, therefore, will lead to increased sympathetic drive and the possibility of hypertensive episodes.

The inhibition lasts for up to two weeks after discontinuation (as new enzyme has to be synthesised) and care must be taken with diet and use of other drugs during this time. Switching to another antidepressant is difficult and an SSRI or tryptophan cannot be started during this time in order to avoid the serotonin syndrome.

Reversible Inhibitors of Monoamine Oxidase A

Moclobemide, a reversible inhibitor of MAO-A (RIMA), specifically and reversibly inhibits MAO-A in therapeutic doses. As the inhibition is reversible, if there is excessive tyramine intake, the tyramine can displace the drug from the enzyme and be metabolised. Dietary restrictions are no longer applicable, although co-administration of drugs such as ephedrine is not advised (BNF 2010). Care should also be taken when switching from another antidepressant but, due to its short half-life, moclobemide can be stopped and another drug started the following day.

Befloxatone is an oxazolidinone derivative which selectively and competively inhibits MAO-A. Its activity is much greater than that of moclobemide and animal studies showed it had promise as an antidepressant (Curet *et al.* 1998). However, it seems to have been dropped from the development pipeline.

Selective Serotonin Reuptake Inhibitors (SSRIs)

Clomipramine was noted to have relatively high affinity for the serotonin transporter, which led to the development of the SSRIs (Nutt, 2002). Although the selective serotonin reuptake inhibitors (SSRIs) have no common structure and are chemically unrelated to each other, they all block the reuptake mechanism for serotonin. Zimeldine, the first SSRI to be commercially available in the UK, was marketed in the early eighties and was a very effective antidepressant but was withdrawn due to serious side effects (Anon 1983). Paroxetine and sertraline are the most potent inhibitors of the serotonin transporter but it is citalopram that is the most selective. Fluoxetine is the least selective. Although no more effective than the TCAs (Guze 1996), SSRIs have proved to be useful agents, increasingly used as first-line drugs. However, some psychiatrists feel that they are not as effective in resistant depression as TCAs (Anderson & Tomenson 1994). They are useful in primary care where they can be used in a therapeutic dose from the start of treatment. The SSRIs are regarded as having a generally cleaner side-effect profile than the predecessors, with less cardiovascular and antimuscarinic side-effects due to the lower affinity they have for muscarinic (M_1), histaminic (H_1) and noradrenergic (α_1) receptors, but they do have the disadvantages of causing nausea and vomiting, headache, sexual dysfunction, and anxiety.

Not all the SSRIs are the same, as although a patient may not respond to one of the class, they may respond to another. This may be due to the binding affinities to and inhibition of the serotonin transporter being different so that the SSRIs are not homogeneous as a class (Nurnberg *et al.* 1999). If reuptake inhibition profiles are considered, only citalopram has the most specific serotonin reuptake inhibition, the others having some inhibition of norepinephrine reuptake, although to a lesser extent.

Escitalopram, the latest SSRI to be marketed, was claimed to have a rapid onset of action and have better efficacy and tolerability than other antidepressants, including citalopram. This drug will be discussed further in chapters 4 and 5.

Noradrenergic and Specific Serotonin Antidepressant

Discovered in the 1970s from a random screening programme in animal behavioural paradigms, mianserin is a tetracyclic antidepressant chemically related to the tricyclic antidepressants (Nutt 2002). It facilitates noradrenegic neurons by blocking an inhibitory presynaptic α_2 receptor. These receptors control monoamine release from terminals via negative feedback. It is now little used because it appears to offer no advantage over the tricyclic compounds and suffers from a potential to cause blood dyscrasias. It is recommended that patients receiving mianserin should have a blood count every four weeks during the first three months of treatment. A further full blood count should be obtained if signs of infection occur during treatment and the drug stopped (BNF 2010).

A similar warning applies to mirtazapine, a noradrenergic and specific serotonergic antidepressant (NaSSa) developed from mianserin. Mirtazapine is a 6-aza analogue of mianserin that is more specific for presynaptic α_2 adrenoreceptors (de Boer 1996), which enhances noradrenergic transmission. It has no effect on reuptake. Serotonin levels rise due to α_1 receptor mediated enhancement of serotonin cell firing, and by blocking the inhibitory α_2 heteroreceptors on 5HT terminals (Richou *et al.* 1995). Mirtazapine not only increases NA and 5HT transmission but also blocks post-synaptic $5HT_{2A}$ and $5HT_3$ receptors which increases the amount of available neurotransmitter in the synapse and allows more serotonin to act at the $5HT_1$ receptor, thought to be important in depression (Davis & Wilde 1996). This dual action may mean mirtazapine has better efficacy in refractory depression. Blockade of $5HT_2$ and $5HT_3$ receptors has been associated with promotion of deep sleep and possibly anxiolytic effects. Stimulation of these receptors leads to sleep disturbances, sexual dysfunction and gastrointestinal stimulation, common side effects of SSRIs.

Mirtazapine lacks the side-chain found in TCAs thought to cause the anticholinergic side effects, so the profile is different to that of the TCAs.

Noradrenaline Reuptake Inhibitor (NARI)

Reboxetine is unique and specifically inhibits NA reuptake (NARI), although it has some structural similarities to fluoxetine (Dostert *et al.* 1997). Specificity for the reuptake mechanism is high, and reboxetine has little affinity for other receptor sites, giving it a clean pharmacological profile (Brunello 1998). It is also claimed to have efficacy in patients who have poor social functioning (Dubini *et al.* 1997; Montgomery 1997b). In his review of four placebo-controlled trials and three active comparator trials, Montgomery

(1997a) concluded that reboxetine was at least as effective as imipramine, desipramine and fluoxetine. Reboxetine would be a useful addition for the treatment of depression, particularly as a subset of patients in the fluoxetine study showed greater improvement in social functioning with reboxetine. However, the social functioning instrument used, the Social Adaptation and Self-assessment Scale (SASS) (Bosc *et al.* 1997), appears to have been developed specifically for testing reboxetine. Another review was more cautious in its conclusions, suggesting that it only may improve social functioning and that well-conducted trials were required to establish its place in therapy (Holm & Spencer 1999). There appear to have been no major trials since launch, although it has been successfully used in bulimia nervosa (Fassino *et al.* 2004), drug addiction (Szerman *et al.* 2005) and in post-stroke depression (Rampello *et al.* 2005).

Serotonin & Noradrenaline Reuptake Inhibitor (SNRI)

This class of antidepressant now has three examples in clinical use: duloxetine, milnacipran and venlafaxine. Venlafaxine (Fig 3) has an interesting pharmacology compared to the others of this class. At doses below 200mg daily, it specifically inhibits serotonin reuptake, while at doses above this it blocks reuptake of both noradrenaline and serotonin (Nutt 2002). At very high doses, it also blocks dopamine reuptake (Mendelwicz 1995). At moderate doses it could be seen as a cleaner version of the TCA, clomipramine, as it has fewer anticholinergic side effects and a much weaker binding affinity profile at other receptor sites (Mendelwicz 1995). This variability in dosing does have the associated problem that doses are usually titrated up until a treatment effect is seen. Although flexible, this can mean more physician visits to alter the dose: some patients may be less compliant with this.

There is also a possibility that venlafaxine has a more rapid onset of action than other antidepressants (Rickels *et al.* 1995; Montgomery 1995). This may be due to its dual action on reuptake systems, particularly if the dose is increased rapidly within one week (Guelfi *et al.* 1995). One study has indicated that venlafaxine may be useful in patients after previous treatment failure (Nierenberg *et al.* 1994), although the study was an open, uncontrolled design, resembling actual clinical practice.

There are indications that milnacipran (Fig 5) may have a more rapid onset of action and is well-tolerated, superior to the SSRIs in severe depression (Tajima 2002; Bisserbe 2002; Puech *et al.* 1997). It is more balanced for inhibition of reuptake or serotonin and noradrenaline, with no inhibition of dopamine reuptake. Milnacipran also has no affinity

for other receptors and is metabolised by conjugation rather than by the CYP450 liver enzymes. These metabolites are inactive, which may avoid the interactions that may be encountered with the SSRIs. Milnacipran is not available in the UK.

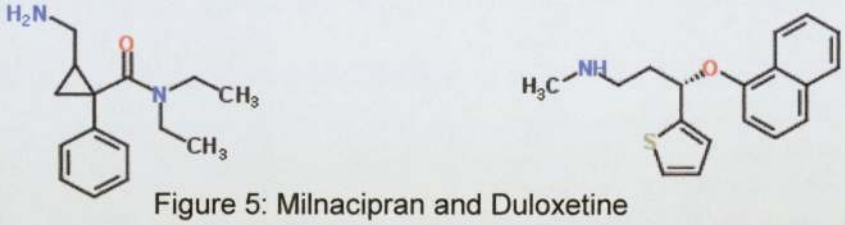


Figure 5: Milnacipran and Duloxetine

Duloxetine (Fig 5) also shows a more balanced inhibition of the serotonin and noradrenaline transporters (Table 2) than does venlafaxine, which indicates a more 'dual action' compound than occurs with venlafaxine (Bymaster *et al.* 2001). This should mean greater efficacy with duloxetine. There is a potentially low risk for adverse effects as duloxetine has relatively low affinity for muscarinic, histaminic, and α_1 adrenergic receptors (Bymaster *et al.* 2001). Short-term safety in the general adult population appears to be good but longer trials need to be conducted to ensure long-term safety and tolerability, particularly in patients with coexisting illnesses (Hudson *et al.* 2004). Six randomised, placebo-controlled acute trials and one 52-week open-label trial have indicated that duloxetine is well tolerated and effective. As it is thought that dual action antidepressants may act faster than those with a single mode of action, duloxetine, like venlafaxine and milnacipran, may have a faster onset of action although a pooled analysis of two trials failed to demonstrate this conclusively (Brannan *et al.* 2005).

Serotonin Antagonist & Reuptake Inhibitors (SARI)

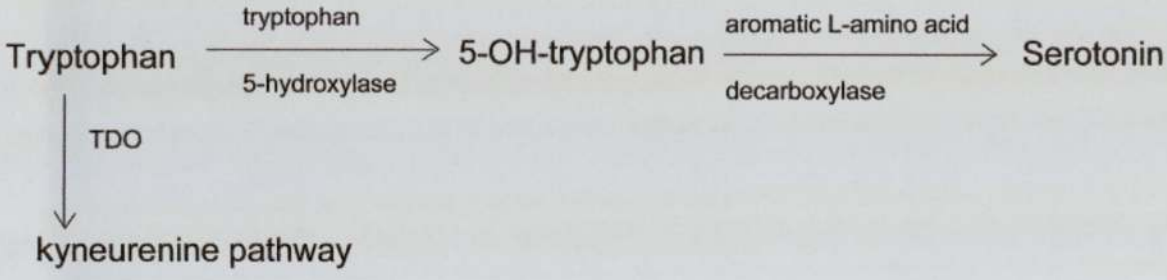
The first SARI, trazodone, only inhibits one transport mechanism; serotonin. Nefazodone, however, inhibits both serotonin and noradrenaline reuptake. Both compounds are potent blockers of postsynaptic $5HT_{2A}$ receptors. The α_1 antagonism of nefazodone tends to be countered by its noradrenaline reuptake inhibition such that little α_1 antagonism results. As trazodone lacks the NA reuptake inhibition, $\alpha_1$antagonism is stronger resulting in more side-effects. It also has antihistamine properties which nefazodone does not.

Blocking 5HT-2 receptors helps reduce anxiety, possibly more quickly than other antidepressants (Mendels *et al.* 1995), enhances sleep, avoids sexual dysfunction, and causes sedation. Nefazodone was withdrawn in 2003 due to hepatotoxicity.

Tryptophan

A more fundamental approach is to increase serotonin concentrations in the nerve terminal, making more available for release. Although tryptophan itself has been used for many years its efficacy is limited because of its rapid metabolism. Vesicular storage of serotonin does not appear to increase with increasing amounts of exogenous tryptophan. Two novel compounds that inhibit the enzyme responsible for metabolising tryptophan, tryptophan 2,3-dioxygenase (TDO) were considered for their potential as antidepressant agents (Salter 1996). Tryptophan can be a substrate for two enzymes: TDO or tryptophan 5-hydroxylase (Figure 6). The major controlling enzyme in the kyneurenine pathway is TDO and is the first enzyme in the pathway. Inhibition of TDO should, therefore, decrease

Figure 6: Tryptophan metabolic pathway



metabolism of systemic tryptophan and thus increase brain tryptophan levels. As tryptophan hydroxylase, the major controlling enzyme of tryptophan synthesis, is normally unsaturated, the extra substrate available to it causes an increase in brain 5HT and its vesicular storage. However, this line of research seems to have foundered, as a search of online databases provided no hits.

Tryptophan is now restricted to use by hospital specialists for patients with severe and disabling depressive illness that has lasted continuously for more than two years, and only after adequate trials of antidepressants. Then tryptophan should only be used as an adjunct. Patients and prescribers have to be registered with a monitoring service, as tryptophan has been associated with eosinophilia-myalgia syndrome.

Adverse Effects

All drugs can cause adverse effects that may be unpleasant enough to stop patients taking their medication. In mental illness, the results of failing to take prescribed medication can be catastrophic. Table 2 describes the binding affinities for the various common antidepressants and predicts the potential side effects. The TCA side effect profile includes including drowsiness, dry mouth, blurred vision, urinary hesitancy, constipation, and hypotension as the most common problems. In overdose, most of the TCAs are particularly dangerous as they are cardiotoxic. Drowsiness and hypotension can impair the ability to perform various psychomotor tasks and cause falls in the elderly. Reaction times are also decreased, an important factor for young active patients in particular.

The dry mouth, constipation, blurred vision, urinary hesitancy and drowsiness are due to antimuscarinic actions at the M_1 muscarinic receptor: drowsiness and day-time sedation are also mediated by antihistaminic effects at the H_1 histamine receptor. Indeed, amitriptyline may be one of the most potent antihistamines, along with doxepin (Richelson 1979). Hypotension is a result of α_1-adrenoceptor blockade. These effects vary in intensity between compounds depending on their binding affinity to receptors. The TCAs also have Class I antiarrhythmic properties, which is the reason for their cardiotoxicity in overdose. An exception is lofepramine, which has minimal cardiotoxicity in overdose. However, it can have adverse effects on the liver, although these are not common.

SSRIs have different affinity profiles and therefore tend to be better tolerated than TCAs and MAOIs and do not have the interactions that the latter have. Discontinuation rates resulting from adverse events with fluoxetine, for example, tend to be less than those with TCAs (Pande & Sayler 1993). One of the major 'selling points' of SSRIs has been the improved side-effect profile and lack of toxicity in overdose. In particular the lack of drowsiness, lack of impairment of concentration and lack of effect on reaction time have been put forward as essential points in their favour. The inference is that fewer accidents will occur as a result of prescribing SSRIs than if TCAs are used instead.

However the activity of serotonin at $5HT_2$ and $5HT_3$ receptors can cause headache, nausea, sexual dysfunction and anxiety early in therapy. These effects can be particularly distressing after initiation and have implications for continuing compliance. Although they are less toxic than TCAs and relatively safe in overdose, and thus of little use to patients intent on suicide, this will not necessarily stop individuals from trying. Some early evidence relating to fluoxetine suggested that it increased the incidence of suicide and this perception has continued. This increase in suicidality may have been due simply to increased motivation in severely depressed patients without the mood lifting sufficiently

(Müller-Oerlinghausen & Berghöfer 1999), and meta-analyses of RCTs have revealed excess suicidal ideation on active treatments compared with placebo, with an odds ratio of 2.4 (95% CI = 1.6 to 3.7) (Healy & Whitaker 2003). The apparent increase could also be as a result of selection bias, in that doctors are perhaps recognising depression more efficiently and also prescribing 'safer' drugs (i.e. the SSRIs) for suicidal patients. However, an increased risk of suicide during the early stages of treatment in patients taking TCAs has long been known (Damluji & Ferguson 1988). Others dispute the idea that antidepressants, particularly SSRIs, can increase suicidal thinking or behaviour (Rich 1999; Lapierre 2003). It does appear that opinion may be divided on this aspect.

Generally, SSRIs are better tolerated but there are no major differences between TCAs and the newer compounds in terms of efficacy or speed of onset. Indeed, withdrawal rates in a meta-analysis of 42 published randomised controlled trials showed that SSRIs are better tolerated, although no significant difference in efficacy between the two groups was seen (Montgomery *et al.* 1994). Combinations of SSRIs and TCAs have been found to produce greater efficacy or slightly faster onset of action. Venlafaxine (Montgomery 1995, Benkert *et al.* 1996), mirtazapine (Claghorn & Lesem 1995; Kasper 1995), milnacipran (Clerc *et al.* 2001; Kasper *et al.* 1996), reboxetine (Montgomery 1997b), paroxetine (De Wilde *et al.* 1993) and escitalopram (Montgomery *et al.* 2001) may possibly have faster onset but there is still no unequivocal evidence for this, so the acceptability of the side-effect profile and the sequelae from overdose become more important considerations in the financial equation (Hale 1994). Venlafaxine is perhaps better tolerated, unless rapid dose escalation is used, and being a dual-action drug (norepinephrine and serotonin reuptake inhibition), potentially offers high efficacy (Clerc *et al.* 1994) but appears to have an increased incidence of cardiovascular adverse events (Combes *et al.* 2001; MHRA 2004). Discontinuation syndrome, although not strictly speaking a side effect, can be a problem when an antidepressant is stopped abruptly due to side effects, lack of efficacy, or non-compliance by the patient. It is not a withdrawal syndrome that implies a potential for addiction; a point to be stressed to patients and carers. The syndrome came to prominence with paroxetine but can occur with other antidepressants (Lejoyeux & Adès 1997). To minimize the symptoms of discontinuation, antidepressants with short half-lives must be gradually tapered. Fluoxetine is the exception presumably due to its long half-life.

Discussion

Western medicine has progressed a long way since Hippocrates and physicians of the first half of the second millennium when depression was thought to be due to evil

humours or black bile. Since the 1950s and 1960s when the monoamine hypothesis was postulated, much has happened in the understanding of how depression may develop. Through imaging studies, it has been shown that there is loss of neuronal tissue in regions of the brains of patients with depression that are important for maintaining mood. This appears to be the result of prolonged hyperactivity of the HPA, the raised corticosteroids from which cause damage to neuronal structure and loss of tissue. The simple idea of depression being caused by a reduction in monoamine activity at post-synaptic receptor sites is only part of the probable truth. When an antidepressant is administered, the drug action is probably only the beginning of a complex train of events in neurons. These events appear to culminate in the expression of BDNF, which enhances the repair and development of neurons, so reconstructing the integrity of brain structures. However, other processes are involved: increased levels of SP appear to be correlated with depression and the discovery that SP NK1 receptors antagonists can relieve depression may lead to a new method of treating the disease (Bondy 2003). It has also been suggested that if stressors can adversely change the connection strength of neurons, then it may also be possible to undo those changes by non-chemical means; i.e. psychotherapy (Jeffrey & Reid 1997).

Animal experiments and clinical studies of depressed patients have helped to elucidate some of the mechanisms involved with the development of depression. A new hypothesis has emerged from these studies that encompasses cellular and molecular events, primarily in the hippocampus. The dynamic nature of the hippocampus, ie its neural plasticity, plays a major role. The indications from these studies suggest that neuronal atrophy and death in the hippocampus, as well as other regions of the brain associated with mood, such as the prefrontal cortex, possibly contribute to the pathophysiology of depression. However, it is unknown if the changes in function and structure of the hippocampus can be modified or even prevented by using antidepressants (Campbell & Macqueen 2004).

Genetic variations in the expression of genes coding for essential components of the post-synaptic signal transduction process may account for the variation in vulnerability of individuals to suffer from the illness, and to what extent. These variations are not necessarily, in themselves, enough to lead to disease but may be affected by environmental factors that are neuronal insults.

It is known that chronic antidepressant treatment (including electroconvulsive shock), treatment with substance P inhibitors or glucocorticoid inhibitors relieve depressive symptoms apparently at a similar rate to the upregulation of BDNF and CREB, which protect neurons from further damage. More brain imaging and post-mortem studies

are needed to confirm such findings. If these mechanisms are so important, they make potential targets for drug therapies.

Depression is a complex mixture of genetics, environmental factors, biochemistry and physiology. Although much of the data is from animal experiments, some of it is being supported by human imaging techniques and, as time progresses and these techniques improve, so will our understanding of this serious mental illness.

There is no drug yet which can treat the illness quickly and effectively, although the pharmacological action of the drugs at their receptor sites is rapid. The basic mechanism has traditionally been formed on the premise that there are low levels of monoamine neurotransmitters in the limbic system, particularly the hippocampus. By preventing homeostatic systems (ie reuptake, metabolism, and negative feedback loops via auto- and heteroreceptors) reducing the levels in the synapse, the available drugs increase the amounts of noradrenaline and serotonin in the system. Serotonin and NE depletion studies support this hypothesis. The pharmacological effect is almost immediate but there is usually a lag time before clinical effect is seen in patients. However, there is also the strong possibility that post-synaptic signal transduction pathways may be disrupted in depression and it is these systems which require time to recover.

The nomenclature for the tricyclic antidepressants was based on their chemical structure but subsequent compounds have been classified according to their pharmacology. However it could be argued that this is only a biochemical or pharmacological convenience developed over time by researchers, or possibly devised by marketing to promote a drug as being different from others already established. Although it is convenient (particularly for industry when marketing new products) to refer to SSRIs or SNRIs as though they are distinct entities, one could argue that they are all monoamine reuptake inhibitors and that these are merely subgroupings which describe the particular neurotransmitters systems on which they act. As it inhibits the reuptake of both norepinephrine and serotonin, venlafaxine might be better described as a MARI, albeit a cleaner one (if by MARI it is taken to mean, conventionally, the tricyclics). The 'tricyclics' as a group include chemically non-tricyclic compounds, so this name should have been long since abandoned. Further, if reuptake selectivity is taken into account, many tricyclics can be considered as serotonin reuptake inhibitors (Li Wan Po 1999).

There has been a perception that SSRIs are not as effective as TCAs, particularly those with high specificity for serotonin reuptake inhibition: ie have little or no dual action. In a review of RCTs comparing TCAs and SSRIs, both as general groups and as subgroups of TCAs (those with a balance NE/5HT or those NE specific), most showed comparable

efficacy (Burke 2004). Where trials claimed superiority of one over another, the difference was usually only one rating point on the Hamilton Depression Rating Scale. This effect difference is too small to be of clinical significance; a difference of at least three points is more clinically relevant (Montgomery 1994).

The evidence, therefore, cannot really say whether one type of antidepressant is better than another (ie dual action or single). It has been suggested that a trial comparing venlafaxine with an SSRI should be performed, with well-defined criteria including adequate doses of venlafaxine (Discussion 2004).

Is there then a way of describing these drugs without recourse to biochemical action? The obvious alternative is by chemical grouping but this would be unwieldy in view of the disparate chemical structures. The problem is how we can reconcile the same mode of action with such different chemical compounds. This may be becoming clearer with the results from animal experiments demonstrating a post-synaptic intracellular transduction cascade (Duman *et al.* 1997; Nibuya *et al.* 1996). Is there even a need for grouping them together in some way when all we really need to know is that they are effective or not? Knowing the mechanism of action is interesting and can be useful from a more academic point-of-view but is it necessary for treating patients? Probably not in general terms, but it can be helpful for eliminating drug classes already tried. It can also help develop augmentation strategies in patients with resistant depression. From a clinician's or mental health pharmacist's viewpoint, when there is a patient to treat, it is more important to know that there is an drug available which has been adequately trialled with appropriate outcome measures.

The latest antidepressant to reach the market (in 2004), escitalopram, was hailed as an improvement on its racemic predecessor, citalopram, in terms of both efficacy and tolerability. The pharmacology is interesting and may well confer advantage over citalopram and possibly other SSRIs. As will be discussed in Chapter 5, escitalopram is highly selective for the serotonin transporter, but does that confer better efficacy or tolerability as suggested by the company, Lundbeck? If there is an advantage, perhaps it is not necessary to have dual action as some proponents suggest. Will the greater cost of escitalopram over the generic citalopram be offset by greater efficacy?

CHAPTER 3

RATING SCALES IN DEPRESSION

Introduction

Depression has been very difficult to define as patients can report many differing symptoms. As long ago as the 1960s, it was noted that patients would report a large range of symptoms (Watts 1966 – 71 symptoms in a sample of 590 depressives (cited in Moran & Lambert 1983)). With so many symptoms describing or being reported in depressive illness, it can be difficult to formulate the precise item descriptions that should be used in a rating scale to precisely define the illness and that can be consistently measured across a diverse population. Indeed, patients may be describing the same symptomatology but in different ways, making the differentiation of symptoms more difficult. Montgomery and Åsberg (1979) took a comprehensive psychopathology scale of 65 items and used an arbitrary cut-off point of 70% to identify the 17 most common items that described depressive illness. Further estimates of sensitivity reduced these to the 10 items that were subsequently used to produce the Montgomery-Åsberg depression rating scale.

Efficacy may be considered the primary criterion by which any drug is judged for it to be launched onto the market, although adverse reactions and toxicity must also be taken into account during clinical trials and may determine a drug's fate. Post-marketing surveillance will bring to light any less frequent adverse reactions.

Efficacy in depression can only be demonstrated by a change in depressed mood, which is much more subjective than the measurements found in other areas of medical research in which clinical response may be measured objectively by changes in physical parameters. In trials of antidepressants rating scales have to be used to assess the depth or severity of illness. The rating scales detecting this change must be reliable, valid and sensitive to that change and must have the ability to detect differences between the drugs under test and (perhaps) placebo. There is also a need to know that any scale chosen actually measures what it purports to measure and is able to do so over time, tracking changes in depressive mood which are clinically meaningful; ie demonstrates reliability and validity. Rating scales quantify the level or severity of illness by measuring the degree of a prognostic indicator of a patient's depression. These are founded on clinical observations. The individual scores are summed to give an overall score which is measured against a cut-off point score on the scale which determines the presence, absence or degree of depression (mild, moderate, severe). So this presents as a nominal system of scoring rather than graded, where the scoring would be on a spectrum of severity. A problem emerges when a patient is only one or two points more or less than the cut-off score, as it

becomes more difficult to decide either how depressed they remain or if they are responding or remitting.

Randomised controlled trials will usually include either the Hamilton Depression Rating Scale or the Montgomery-Åsberg Depression Rating Scale (sometimes both) as the assessment scale. Occasionally self-rating scales have been used as well but this is less frequent. All trials will also include a confirmation of depression by using a diagnostic classification system and Clinical Global Impression (Guy, 1976).

The objective of this chapter will be to assess the rationale for using particular depression rating scales that are or have been employed in RCTs. As the data from such trials provide the basis on which to evaluate new antidepressants, either quantitatively or qualitatively, it is essential to know that that data are derived from sound assessment tools. This will support the analysis of the data from trials of escitalopram in Chapter 5. Therefore a systematic review each of the main rating scales, testing for evidence of validity and reliability, is presented. The difference between rating scales and diagnostic criteria is also described, and the significance and utility of the Clinical Global Impression is considered.

Method

Copies of the diagnostic classification systems were obtained and a search for references discussing them was made. A systematic search of papers since 1960 was carried out, looking for those concerning the commonly used rating scales of today for assessing severity of depression in the adult population only and used in randomised clinical trials. The time-point of 1960 was chosen as this was the year that Hamilton published his report on a scale that has since become the gold standard. MEDLINE, BIDS and hand-searching were used to find papers using the keywords depression, rating scale, Hamilton, Montgomery, Åsberg, MADRS, Zung, Beck. Older scales no longer in common usage were discarded. Because trials designed for obtaining a licence do not usually include children or older adults, scales for these groups were not included in the current analysis as the focus for this study was on scales specifically used in clinical trials and used for original licence submissions. The original papers were reviewed qualitatively for their descriptive content. Both psychiatrist-rated and self-rated scales were included, although few RCTs employ self-rating scales. A search for commentaries on the scales was also performed. Keywords included depression, rating scale, Hamilton, Montgomery, Åsberg, self-rated, self-rating, Beck, Zung.

Each scale was inspected for the methodology of its construction, the type of statistical analysis used and the application of reliability and validity criteria used to verify the scales' accuracy in tapping the symptoms of depression. The scales reviewed are listed in table 3.

RESULTS

Diagnostic Classification

Coding systems provide a classification of an illness: useful for demographics, costing analyses, and to enable uniformity in studies. A coding system is a list of the signs and symptoms of the illness, with no rating of their severity. Rating scales go further by describing the severity or intensity of that illness and providing a means of tracking progress of the patient when treated for it. It is important that the two should be correlated, one mapping onto the other to ensure the rating scale is measuring the diagnosed illness. Table 3 describes the items contained within diagnostic systems and rating scales. For the latter, a percentage weighting is given for each symptom as rated within that scale. Some scales do not rate all items as classified in diagnostic criteria and have other descriptors. Larger weights include multiple items in that scale describing that symptom.

The two principal coding systems are the Diagnostic and Statistical Manual (American Psychiatric Association, 1994) & the International Classification of Diseases (WHO 1992). These coding systems rely on a number of symptoms being present to form a diagnosis. They were developed to enable clinicians to describe an illness according to observed symptoms. First developed in the early 1950s as a variant of the ICD-6, the American Psychiatric Association's *Diagnostic and Statistical Manual* has gained ground as the standard glossary of diagnostic criteria for describing a disease, particularly in randomised controlled trials. It is multiaxial, in that it describes operational criteria stating which symptoms need to be present, as well as exclusion criteria. Unfortunately, the early versions did not contain explicit criteria for psychiatric diagnoses and clinicians or researchers were forced to select the diagnostic category closest to the patient characteristics. The inclusion of the term 'reaction' in DSM-I was influenced by Adolf Meyer's psychobiological view of mental disorders, which said that they were personality reactions to psychobiological, social, and biological factors. This definition of illness improved with successive manuals, but did lead researchers to create their own diagnostic classifications such as the Research Diagnostic Criteria (Spitzer 1978). However, publication of DSM-III in 1980 saw the introduction of explicit diagnostic criteria, a multiaxial system, with a descriptive approach that was not influenced by aetiological

theories and covered the affective, cognitive, behavioural and physiological features of depression (Moran & Lambert 1983). Table 3 lists the criteria for diagnosis according to DSM-IV, which has refined some of the criteria in DSM-III (and its revision).

The International Classification of Diseases (ICD) was formulated in 1899 by the International Statistical Institute as a system for classifying causes of death. In 1948, under the auspices of the World Health Organisation, the sixth edition of ICD was published as a basis for mortality statistics and included for the first time a section on mental disorders. The ICD is now in its tenth edition. Table 3 highlights the differences between this system and the DSM. ICD is a uniaxial, hierarchical system because it uses descriptive terms of the illness as well as having directives on differential diagnosis. It is used as a standard coding system for describing medical care in both the USA and the United Kingdom. In the USA, it is commonly used for reimbursement of care and because of this use for capturing clinical data, it is useful for health care research for costs and outcomes.

The Research Diagnostic Criteria formulated by Spitzer (Spitzer *et al.* 1978) was a response to what was seen as a failure of existing systems to reliably diagnose and classify a mental disorder. Prior to this, researchers had had to develop their own explicit criteria and classification systems. This could have led to difficulties in comparing trial reports for drugs if a different system was used in each report. Spitzer attempted to standardize the nomenclature of mental disorder by modifying criteria that had been developed by Feighner some years earlier (Feighner *et al.* 1972), improving on some of the definitions of the earlier work. Some of these revisions were incorporated into DSM and ICD: for example, the criterion of having low mood for two weeks for a definite diagnosis. In the

Table 3 Comparison of Diagnostic Systems & Rating Scales

	Diagnostic Systems					Rating Scales						
	RDC	ICD-10	DSM-III	DSM-IV	Feighner	HAMD	MADRS	BDI	CRS	EPDS	HADS	ZUNG
No of Items	12	10	13	11	13	17	10	21	17	10	7	20
Max Total Score						50	60	62	52	30	7	80
Symptom						*Weighting expressed as percentages*						
Depressed mood	#	#(am)	#	#	#	5.88	20.00	9.52	9.6	20.00	14.28	10.00
Anhedonia	#	#	#	#	#	5.88	10.00	9.52		20.00	71.42	15.00
Low self-esteem	#		#	#	#			14.28	3.85			5.00
Guilt	#		#	#	#	5.88	10.00	9.52	3.85	10.00		
Fatigue	#		#	#	#			4.76	5.76		14.28	5.00
Retardation	#	#	#	#	#	5.55	10.00	4.76	7.70	10.00		5.00
Suicidal thoughts & action	#		#	#	#	5.88	10.00	4.76	7.70	10.00		5.00
Appetite	#	#	#		#		10.00	4.76	3.85			5.00
Weight	#	#	#	#	#	5.88		4.76	3.85			5.00
Libido		#	#		#	5.88		4.76	3.85			5.00
Sleep	#	#(am)	#	#	#	17.64	10.00	4.76	11.53	10.00		5.00
Somatic Symptoms						17.64			11.55			10.00
Psychic anxiety					#	5.88	10.00		3.85	20.00		
Agitation	#	#	#	#	#	5.88		4.76	7.70			5.00
Concentration	#		#	#			10.00		1.92			
Irritability								4.76	1.92			5.00
Loss of insight									3.85			
Hypochondriasis								4.76				
Somatic preoccupation								4.76				
Body image								4.76				
social withdrawal									1.92			
Meeting people									3.85			
Somatic concern									1.92			
Life worth living								4.76				5.00
Indecisive												5.00
Mind clear												5.00
Diurnal rhythm												

Key:
RDC: Research Diagnostic Criteria
ICD: International Classification of Diseases
DSM: Diagnostic & Statistical Manual
Feighner:
HAMD: Hamilton Depression Scale
MADRS: Montgomery-Asberg Depression Rating Scale
BDI: Beck Depression Inventory
EPDS: Edinburgh Post-natal Depression Scale
CRS: Carroll Rating Scale
HADS: Hospital Anxiety & Depression Scale
ZUNG: Zung Self-rating Scale

original Feighner criteria, this was a month. Reliability was tested in three studies, the first of which used an early draft of the RDC, while the last two used the first and second editions. The first two studies used pairs of raters, while the third used a test-retest method. A total of 278 patients were assessed and reliability was found to be high, with kappa coefficients of agreement being in the order of 0.9. Clinical Global Impression (Guy 1976) is commonly used in clinical trials for depression. It consists of three assessments, although only two are used in randomised controlled trials. The first assesses overall severity of illness (from 'normal, not ill' to 'among the most extremely ill patients'), used at initial and subsequent assessments. The other two parts do not have to be done at the initial assessment but are used for subsequent interviews. One is a measure of global improvement, while the other provides an index of efficacy. Global improvement rates the improvement in the patient's illness whether or not it is entirely due to drug treatment: this can be from 'very much improved' to 'very much worse'. The efficacy index attempts to tease out the proportion of improvement due to drug effect, weighing efficacy vs. side effects. It does not appear to be used in trials.

The Scales

Twelve scales were identified (Beck 1961; Bech & Rafaelson 1980; Bech *et al.* 1997; Carroll 1981; Cox 1987; Hamilton 1960; Montgomery & Åsberg 1979; Ottosson 1960; Rush 1996; Snaith *et al.* 1976; Zigmond & Snaith 1983; Zung 1965). Five were discarded because they are now rarely or not used, particularly in RCTs (Bech & Rafaelson 1980; Bech *et al.* 1997; Ottosson 1960; Rush 1996; Snaith *et al.* 1976). Of the remaining seven scales (Table 4), 3 are physician rated (Hamilton 1960; Beck 1961; Montgomery & Åsberg 1979) 3 are self-rating (Carroll 1981; Zung 1965; Zigmond & Snaith 1983), while the Edinburgh post-natal (Cox 1987) is designed to be used by untrained raters or by patients. Only two of these scales, the Hamilton Depression Rating Scale and the Montgomery-Åsberg Depression Scale are widely used in randomised controlled trials. Self-rating scales have sometimes been employed and, when they have been, tend to be either the Beck Depression Inventory or the Zung, although use of the latter has fallen to almost, if not completely, zero in clinical trials. The Beck appears to be used more in clinical psychology settings.

Table 4: Common rating scales

Author	Scale Name	Year Published	Derivation
Beck	Beck Depression Inventory (BDI)		Empirical
Carroll	Carroll Depression Rating Scale (CRS)	1981	Based on Hamilton scale
Cox	Edinburgh Post Natal Depression Scale (EPNDS)	1987	IDA, HADS, ADS [1]
Hamilton	Hamilton Depression Rating Scale (HAMD)	1960 & 1967	Empirical
Montgomery Åsberg	Montgomery-Asberg Depression Rating Scale (MADRS)	1979	CPRS[2]
Zigmond & Snaith	Hospital Anxiety & Depression Scale (HADS)	1983	Empirical
Zung	Self-Rating Depression Scale (SDS)	1965	Empirical

1 IDA: Irritability, Depression & Anxiety Scale. A Clinical Scale for the Self Assessment of Irritability Snaith *et al.*, 1978 Brit J Psych: 132; 164-171
ADS: Anxiety and Depression Scale. Delusions-Symptoms Stakes: State of Anxiety and Depression (Manual). Bedford & Foulds, 1978 Windsor National Foundation for Education Research
2 Comprehensive Psychopathological Rating Scale. Åsberg *et al.*, 1978 Acta Psychiatrica Scandinavica: Suppl 272; 5-27

HAMILTON DEPRESSION RATING SCALE

This scale was developed to overcome several limitations of the then existing rating scales for quantifying the severity of depression in individuals already suffering from depression (Hamilton 1960). The existing scales at that time were not specifically designed for elucidating depressive symptoms. Many were devised using normal subjects, in whom there is no loss of function as there is in illness, while others covered the whole range of possible symptoms, where it can be very difficult to differentiate between symptoms describing depression and those that describe other mental illnesses. The self-rating scales available suffered from 'the notorious unreliability of self-assessment' and were of 'little use for semiliterate patients'. There was also the problem that patients with serious illness were unable to complete such scales. The last group of rating scales assessed behaviour, and social adjustment of patients in hospital wards: they gave little or no symptom description (Hamilton 1960).

The original scale consists of 17 variables, some of which are defined 'in terms of a series of categories of increasing intensity', others by a number of equal-weighted terms. Four additional variables are also included on the form: diurnal variation, de-realisation, paranoid symptoms and obsessional symptoms but these are not included in the main scale. Hamilton also suggested using two raters at an interview and summing the scores. Otherwise one rater doubles the score obtained. To investigate the utility of the scale, Hamilton started with a cohort of 70 patients but only included data from 49 of them for calculating the product-moment correlations. It was considered sufficient that the ratings were 'repeated often enough to make the individual variables highly reliable'.

Hamilton further validated his scale with results using data from 152 male and 120 female patients (Hamilton 1967). At the same time, he also expanded item 9 (agitation) to give a score of 0-4, instead of 0-2 as in the original 1960 version. He found that the new correlation matrix was much as that derived in the original paper. He also separated the scores for men and women to compare the two groups based on his belief that there were differences between the sexes, such as the incidence being greater in women than in men. Hamilton also suggested there are differences in the pattern of symptoms between men and women. He found the males had high scores on items which were different to those for the females. Males rated higher on items 4, 7 and 8 (initial insomnia, inability at work and loss of interest, retardation) than females who rated higher on items 1, 9, 10 and 13 (depression, agitation, psychic anxiety and fatigability). For the remaining 10 items, no significant differences were seen. Several modifications of the HAMD scale are now in use, thereby causing potential confusion in interpretation of results from the different scales (Grundy *et al.* 1994).

To develop the original scale (Hamilton 1960), two raters scored ten patients at a time, and the correlation between summed scores calculated. Adding successively 10 patients at a time, the correlation changed from 0.84 for the first ten to 0.9 for the last 10, which represents the correlation for all 70 patients. This could be an indication of the scale integrity, or it may be the raters were improving and showing convergence in their rating skill.

For the first 49 male patients, product-moment correlations were calculated for the 17 variables. The correlation matrix produced was then factor-analysed by extracting latent roots and vectors to produce six factors, which grouped together the better correlations in each factor. (Factor analysis takes a large set of variables and attempts to reduce them into smaller groups, or 'factors', which contain correlated variables.) Intercorrelations were found to be low due to the 'intense selection' of patients (Hamilton does not describe his patient group, other than they were male). The first four of six factors were used for calculating factor measurements for the patients, in the form of T-scores.

Factors 1 and 2 showed reasonable 'correspondence with the classical descriptions' of 'retarded depression' and similar to agitated depression, although factor 2 had little depressed mood. The third and fourth factors did not match any particular clinical pattern.

Hamilton also reviewed the scores for patients for each factor and found a good correlation between a factor score and the clinical impression. (E.g. high factor 1 score correlated well with endogenous depression.)

Tests of significance were not calculated, on the basis that if a large enough number of patients was rated, even the smallest factor would become statistically significant.

MONTGOMERY-ÅSBERG DEPRESSION RATING SCALE

This scale was developed with a view to improving on the sensitivity of existing scales with respect to change in symptoms over time (Montgomery & Åsberg 1979). This is important for comparing changes in the severity of depression in controlled clinical trials of new antidepressants with active comparators. Therefore the authors constructed a scale which included only items that showed sensitivity to change. This was made possible by rating 106 patients against 65 scaled items of the Comprehensive Psychopathological Rating Scale (CPRS) (Åsberg *et al.* 1978) which covers a wide range of psychiatric symptoms.

The selection of items was based on ratings for 106 patients, 33 men, 73 women who were participating in clinical trials of antidepressants. Two raters were usually involved in interviews. Only patients with primary depressive illness were included. Inventories developed by Gurney *et al.* (1972), were used to ensure diagnostic and descriptive uniformity. Patients were included from two countries to eliminate cultural bias in item selection.

CPRS scores after 4 weeks therapy with four different antidepressants were used to study change with treatment. Scores were available for 64 patients at baseline and 4 weeks. Thirty-five of these patients were simultaneously rated on The Hamilton Depression Rating Scale and on a 7 point scale for global severity of illness.

Parametric statistical methods were used (these are not described) except when analysing ranked data. Due to high agreement between CPRS scores frequencies above zero and ranking by incidence for English and Swedish raters, the two samples were merged for further analysis. The 17 most common items were then identified by using an arbitrary cut-off point of 70% occurrence. The severity of illness as determined by the sum of scores on these 17 items was significantly correlated with both HAMD ($r=0.94$, $p<0.001$) and global scores ($r=0.89$, $p<0.001$) during the fourth treatment week. These correlations were slightly lower before treatment.

Sensitivity was estimated in two ways. First, the mean changes of scores (absolute values) on each of the 17 items after 4 weeks were calculated and ranked. Second, correlation between change on each item and overall change on the preliminary 17-item scale over the 4 weeks were calculated. An item should show a large change (which can be

reliably rated) and be strongly correlated to the general reduction in depressive symptoms. The summed ranks from both estimates were used to select the 10 most sensitive items for the final rating scale.

For inter-rater reliability estimation, data from conjoint interviews were used. Comparisons between 2 English, 2 Swedish and 1 of each were used. Values for r were 0.89 or greater, $p<0.001$. Similar correlations were found for untrained raters when testing for robustness of the scale.

To test validity, comparison was made with global judgement on a sample of patients where there was a clear-cut difference between responders and non-responders and which scale differentiated best. The preliminary 17-item scale, the final 10-item scale and the HAMD were tested by calculating point biserial correlations between response category and change scores. All correlations were highly significant. The 10-item scale was able to discriminate best ($r=0.70$), with the 17-item version and HAMD having correlations of $r=0.67$ and $r=0.59$, respectively.

BECK DEPRESSION INVENTORY

This 21-item inventory (BDI) was published within a year of the HAMD (Beck 1961). Beck *et al.* had been concerned about the lack of agreement on clinical diagnosis of depression. He therefore attempted to formulate a reliable and valid method based on measurement of the behavioural manifestations of depression.

Inventory items were primarily clinically derived by systematic observations of the attitudes and symptoms of depressed patients. Beck then constructed a 21-item symptom and attitude inventory. Each item is scored from 0-3. Originally the score reflected the patient's state at the time of interview but then was changed to reflect the patient's attitude during the previous 7 days. This has now been modified in the updated version (BDI-II) to include the previous 14 days to be more in line with current diagnostic criteria of DSM-IV.

The authors used a 'Depth of Depression' (D of D) global rating to assess the reliability of the psychiatrists' ratings on the depression inventory. This 4-point scale (none, mild, moderate, severe) seems to be similar to Clinical Global Impression (Guy 1976). There was agreement between psychiatrists when using this Depth of Depression measure to within one degree of disparity in 97% of cases indicating a high degree of agreement between raters.

An important aspect of an inventory is its ability to measure change over time. 38 hospitalised patients were retested at an interval of 2-5 weeks. In 5 cases, the D of D category had not changed but there were fine changes in depression severity. In the

remaining 33 patients there were enough gross changes to move from one D of D category to another, but in all cases the inventory scores changed, reflecting its ability to track minor changes. In 28 out of the 33 cases, D of D was predicted by a change in the inventory score.

The inventory was tested in two ways. First, 200 consecutive cases were analysed by comparing the score for each of the 21 categories with the total inventory score for each patient. Kruskal-Wallis Non-Parametric Analysis of Variance by Ranks showed a significant relationship between all categories and the total score. Significance was beyond the 0.001 level for all categories except for weight loss which was significant at the 0.001 level.

The second evaluation of internal consistency was by estimation of the split-half reliability, for which 97 cases in the first sample were selected. The Pearson r between odd and even categories was calculated to be 0.86, a value that increased to 0.93 with a Spearman-Brown correlation.

To test the stability of the inventory, the authors used variations of the test-retest method and the inter-rater reliability method. For the former, at the time of each test administration, a clinical estimate of the Depth of Depression was made by one of the psychiatrists. Changes in inventory score tended to parallel depth score, indicating a close relationship between the patient's clinical state and the depression inventory.

For inter-rater reliability, each inventory score for each of the three interviewers was plotted against the clinical ratings. A very high degree of consistency among interviewers was found at each level of depression.

The inventory was validated in several ways. A comparison of the mean scores for the inventory in each category of Depth of Depression shows an increase of the inventory scores for each increase in the 'magnitude' of depression. Kruskal-Wallis One-Way Analysis of Variance by Ranks showed statistically significant differences, with p-values <0.001, describing an overall association between inventory scores and Depth of Depression ratings. The Mann-Whitney U test was then used to estimate the discriminatory power of the inventory between specific Depth of Depression categories. All differences between adjacent categories in both studies were significant at the 0.0004 level except for the moderate and severe categories, which had a p-value of <0.1 in Study I and <0.2 in Study II.

They also calculated numbers of false positives and false negatives when inventory scores were plotted against Depth of Depression ratings, using non-adjacent D of D categories. Discrimination was seen, particularly in Study II when it was assumed that the

psychiatrists were more experienced. Also greater discrimination is seen with extreme groups (none vs severe). A "cutting score" was utilised as a cut-off point to differentiate between positives and negatives. In Study II, 91% of cases were discriminated in the extreme groups.

CARROLL RATING SCALE

The Carroll Rating Scale (CRS) was designed as an adaptation of the original 17-item HAMD (Carroll *et al.* 1981), to enable self-rating. The scoring method remains the same, with a maximum score of 52. In the Hamilton scale, items are scored either 0-2 or 0-4 and are represented in the CRS as either 2 or 4 statements of increasing severity, each statement scoring one point towards the total. No weighting was applied to the statements, to allow better comparison with the Hamilton score by doctors using the CRS. Patients completed a form in which the statements are presented randomly. To test the scale, a cohort of 119 adults aged 18-64 from the general population completed the CRS, along with over 200 patients being treated for depression. Psychiatrists also rated patients on a global 4-point severity of depression scale. The mean score from the general population was 4.6 (SE 0.4) with the distribution leaning heavily towards low values and with a median of 3. Based on these findings, the authors suggested a cut-off score of 10, above which subjects would be regarded as depressed.

Concurrent validity was estimated by comparing CRS scores with HAMD scores in patients suffering with endogenous depression. Patients were also scored with the Beck Depression Inventory. Correlation and partial correlations were determined for the three scales.

Internal consistency of the CRS was estimated by correlating individual item scores with the total score at the same time as matched HAMD ratings. The split-half reliability of the CRS was tested by correlating the sum of odd- and even-numbered items with each other and the total score. A total of 3725 ratings were available for analysis. Odd sums correlated well with even sums ($r=0.87$, $p<0.001$) and the sum of each half-set was highly correlated with the total score ($r=0.97$ for odd, 0.96 for even). Similarly high correlations were found when looking at the yes and no statements vs the total score.

When examining the total scores for CRS and HAMD, a correlation of $r=0.80$ ($p<0.001$) was found. Matching items on the 2 scales generally correlated although some items did not. Both scales were examined for internal consistency. In both scales, items that correlated strongly with the total were usually also strongly correlated with other items within the same scale; the converse holding true for weakly correlated items. In the CRS,

the median correlation was 0.55 while in the HAMD, individual items had a median of 0.54. Also the rank order of CRS item correlations with the total score was similar to the rank order of HAMD item correlations with total score.

ZUNG SELF-RATING DEPRESSION SCALE

This scale was developed for assessment of depression and sleep disturbances in patients with a primary diagnosis of depressive disorder (Zung 1965). It was designed to be short, simple to complete and self-rated. This 20-item scale was based on the most commonly found characteristics of depression according to the findings of 3 authors. This formed the basis of diagnostic criteria from which a scale was constructed, using patient interviews to find those statements that most represented a particular symptom. The scale was devised so that 10 items were worded symptomatically positive, 10 negative. The scale applies at the time of testing and was constructed so that less depressed patients have low scores, converse for the more depressed patients.

An index was derived by dividing the sum of the scores obtained on the 20 items by the maximum possible score of 80 and expressed as a decimal (1.0 being the maximum). A total of 56 patients was tested over a period of 5 months. Of these 56, 31 were eventually diagnosed as having depressive disorder; the remaining 25 were diagnosed as having other psychiatric disorders. The scale was also given to 100 normal controls.

The authors looked at the physiological and psychological concomitants, grouping them into 'thirds' of worst through least symptoms. They found that patients with other psychiatric illness could present with depressive symptoms (eg poor sleep, irritability) although the final diagnosis is not depression. Other symptoms, they suggest, should perhaps be accounted for when diagnosing depression; eg fatigue, decreased libido, decreased appetite, suicidal tendencies.

This scale appears to have been developed for a specific purpose by the authors who were studying sleep disturbances in depressive disorders, but it does show that poor sleep and irritability do not necessarily equate to a depressive state.

Scores for the depressed group had a mean index of 0.74 before treatment, while the mean index for other disorders was 0.53. After treatment, the mean index dropped to 0.39 in the depressed group. Normal subjects had a mean of 0.33. The only statistics used were t-test: controls vs depressed had $p<0.01$. No significant difference was found between these groups after treatment. No other reliability or validity tests appear to have been used.

EDINBURGH POST-NATAL DEPRESSION SCALE

The authors, Cox, Holden and Sagovsky, suggested that existing instruments for screening for depression were inadequate when used on childbearing women (Cox *et al.* 1987) and that scales such as the BDI could give misleading results. They suggested that the normal physiological changes associated with childbearing may give rise to the somatic symptoms associated with psychiatric disorders. A postnatal depression scale must therefore be appropriate to the situation, short and easy to complete and acceptable to women who probably feel normal.

A detailed analysis of the items found in 3 existing scales was carried out and 21 items developed, including several of the authors' own construction, which were thought appropriate to the detection of postnatal depression. To test these items, extensive pilot interviews were carried out. Thirteen items were selected from the initial 21, seven constructed by the authors.

The authors also found that mothers interviewed with family members present tended to either exaggerate or minimise their symptoms. The highest false positive scores and three of the four false negative scores belonged to subjects who had had another family present at interview.

Validity of this 13-item scale was tested on 63 women, which showed that a clear distinction between depressed and non-depressed women could be made. However, a rotated factor analysis showed that two of the irritability subscale items from one of the selected scales (the Irritability, Depression and Anxiety Scale (Snaith *et al.* 1978)) and an item concerning the enjoyment of motherhood formed a separate 'non-depression' factor. A further study on the 10-item scale formed by removing these items was carried out on 84 mothers, as analysis suggested removal of these items increased the specificity of the scale. Criteria used for diagnosis of a depressive illness were the Research Diagnostic Criteria (RDC) of Spitzer *et al.* (1978). Validation of the 10-item EPDS was determined by comparing the EPDS scores with the RDC clinical diagnosis of depression.

All of the 21 women with an RDC diagnosis of Definite Major Depressive Illness and 2 of 3 women with Probable Major Depressive Illness were identified with a threshold score of 12/13 on the EPDS scale. 86% of women who were true positives and were RDC depressed describes reasonably high sensitivity. The specificity was 78%: the proportion of non-depressed women who were true negatives. The positive predictive value, the proportion of women above the EPDS threshold (n=41) who met RDC criteria for

depression (n=30), was 73%. The authors suggested the cut-off point could be lowered to 9/10, which would reduce the failed detection of cases to fewer than 10%.

The split-half reliability of the scale was 0.88, with a standardised α-coefficient of 0.87.

Further analysis showed that the scale could track changes in mood. Mothers who were RDC depressed at both interviews showed no significant change in EPDS scores, while mothers depressed at first interview, but not at the second (again using RDC criteria), had significant score reductions. (EPDS-1 score = 15.8, EPDS-2 mean score = 9.8, $t = 3.72$, $p = 0.002$.)

HOSPITAL ANXIETY & DEPRESSION SCALE

This scale was developed to detect states of depression and anxiety in medical outpatients (Zigmond & Snaith 1983). It was felt that scores derived from existing depression rating scales are affected by physical illness and there is insufficient distinction between differing mood disorders. The authors aimed to distinguish between the concepts of anxiety and depression using a scale that was easy to complete by patients. The scale was completed by the patient while in the waiting room before their appointment, followed by an assessment by the researchers who had no knowledge of the self-rated score. No details of the interview structure are given. They also requested hospital staff to complete the scale, discarding those who reported being under treatment or thought they needed treatment for any 'nervous disorder'. The results from this part of the study are not reported.

The scale is split into two subscales: depression and anxiety. The seven depression subscale items are largely based on the main psychopathological feature of depression, anhedonia, which responds well to antidepressant treatment. Severity is rated using 4-point scale items.

Data were collected on 50 adult patients of both sexes, aged 16-65. Internal consistency of the two subscales was tested using Spearman correlations. One item (awake before need to) had a weak correlation ($r=0.11$) and was removed. The weakest item on the anxiety subscale was also removed to preserve a balance of items on the two subscales. The remaining items on the depression scale had correlations from 0.6 to 0.3, all significant beyond $p<0.02$.

The first 50 patients were analysed into non-cases, doubtful cases and definite cases with appropriate scores of 0-1, 2, and 3-4 respectively. To test the reliability of these

findings, the next 50 patients were analysed similarly. The results were similar so the subscales were adopted.

These data were then examined to see if these subscale scores could be used to indicate the severity of depression (and anxiety). Spearman correlations were calculated, giving a result for depression of $r=0.70$ ($p<0.001$) with the conclusion that the subscale could be used as a measure of severity.

The authors further investigated whether the subscales differentiated between different aspects of mood disorder or merely represented a general index of 'emotional disturbance'. This was done by selecting patients in whom assessors had recognised a distinct difference between the severity of depression and anxiety. 17 patients were found to have a difference of 2 or more points in severity. Interview ratings correlated well with patient ratings for the appropriate scales but there was insignificant correlation for contrary disorders. The sub-sample was small but is an indicator of a possible trend.

The influence of physical illness was investigated by extracting all data sets with ratings on both scales of 0 and 1, matching for age and sex with the normal sample and the difference tested with Student's t-test. For the depression sub-scale score, $t=0.17$ (not significant), so physically ill patients with no mood disorder had similar scores to the normal sample and therefore physical illness does not affect scale scores.

Reliability and validity

On repeated measurements, a test should yield similar results each time, whether the measurements are made on the same individuals on different occasions or by different observers. If there is no change in the individual, then repeated measurements should give results that show random variation only. The extent to which this occurs describes the reliability of the test. Common reliability tests include test-retest reliability, inter-rater reliability and internal consistency. Test-retest is perhaps more difficult to assess in psychopathology ratings as psychiatric illness can vary over time. Subjects may recall answers if the test interval is too short (i.e. less than two weeks) but the interval must not be too long for the depression to have changed, although this is precisely what is required of a scoring system in RCTs. Inter-rater reliability demonstrates the degree to which different raters will agree on scoring: the coefficient value for reliability here should not fall below 0.70 for results to be trusted. This reliability coefficient is commonly calculated using a Pearson correlation coefficient between the two scores: either test and retest scores or those based on the two observers (Streiner 1993). However, some care is needed to ensure that scores are independent of each other. Also the Pearson coefficient is sensitive

to differences in association but not in agreement; raters may be consistent in their scorings, and in relation to each other, but their scores may not agree with others (Streiner 1993).

Internal consistency estimates the degree to which the scale accurately measures the illness. In some respects, inter-rater reliability and internal consistency are probably more important than test-retest reliability, the first because of the need to ensure the scales can give consistent results when used by different raters. The second because all of the items in the scale should measure the various attributes of the illness to varying degrees. If items on the scale are not answered consistently, this implies the scale is either measuring different things or subjects are not able to answer items in a consistent manner. The commonly used internal consistency measure is Cronbach's α (alpha) which has a range between 0 to 1 (Cronbach 1951). Scales with coefficients less than 0.7 should usually be avoided but the value of alpha usually increases with the number of items in the scale. In addition, raters must be trained to administer the instrument so all raters are scoring to the same degree of accuracy and diagnostic capability.

Face and content validity show the extent to which the test is able to measure that which it is supposed to. Note that an instrument rating depression may be reliable but not valid: results may be consistent but it is not measuring symptoms of depression. That is it could be measuring symptoms common to other illnesses. For face validity the scale must contain items that, on the face of it, appear to be measuring the state of depression, and are asking questions about the affect of the person completing the instrument. If questions are irrelevant or appear irrelevant individuals could well leave some unanswered. Content validity of depression scales takes this further by ensuring the scale is tapping depressive symptoms. An instrument should tap all relevant symptoms of depression while leaving out items that are either unrelated to the illness or equivocal and might be found in other illness states. Split-half reliability and comparison of item scores with the total score will assess the internal validity of the scale. Content validity, that is do the items avoid measures of other disease states, has been estimated by comparing the investigational scale against a validated existing scale (eg the MADRS).

Concurrent criterion validity shows whether a scale correlates well with another scale used as a criterion. This reference scale is usually one chosen as a 'gold standard'. However, there is a danger here. If the new scale correlates too well with the gold standard (0.70 or more) then the new scale is not tapping anything different to the older one and may beg the question as to whether the new scale is needed. Correlations between 0.30 and 0.70 indicate the scales are tapping similar symptoms but the new one is probably

sufficiently different in content to be more valid than the gold standard. Equally it could mean the differences are so great that the new scale is estimating something different and therefore not applicable, particularly if the correlation is below 0.3.

In other areas of medicine, it is possible to have defined criteria that can be measured objectively such as blood pressure and heart rate. In psychiatry, however, instead of such objectively measurable signs, subjective symptoms are usually measured and therefore a construct is used to describe the cluster of symptoms that the affected patients experience. At some point in scale development, construct validity may need testing by making predictions based on the construct using different groups of individuals. This can be tested using factor analysis or other statistical methods. Each study with a positive outcome strengthens the construct and the scale. However, the construct itself may vary according to the approach taken by the investigator and over time. The view of depression has been modified over the last century from time to time.

The two scales used by investigators in clinical trials are either the Hamilton or the Montgomery-Åsberg. The use of other investigator-rated scales instead of either or both of these is infrequent, although self-rating scales may be employed on rare occasions in addition to investigator-rated ones. The other rating scales reviewed here are self-rating scales completed by the individual rather than the observer and appear to be rarely used in clinical trials, possibly due to overvaluation by patients of the severity of their illness, although the Beck Depression Inventory and the Zung have been used occasionally. Investigator-rated scales may have the advantage of impartiality but may suffer from an underestimate of severity.

Hamilton developed his scale for use on patients already diagnosed with depression. He did not describe how the items were developed but the scale appears to have been developed from clinical experience. This contrasts with the development of the MADRS and Beck inventory that were specifically formulated from clinical observation. The scale is also used in situations where the diagnosis is suspected, where the HAMD rating of the patient's severity of illness is used for confirmatory purposes. In RCTs, the scale is used after a diagnosis, which is usually confirmed with DSM-III/IV or ICD-9/10. Hamilton did not give the demographic details of the patients enrolled in his initial and follow-up studies except for gender. Factor analysis was used to identify factors to be included. The scale has high inter-rater reliability (Moran & Lambert 1983) and probably for this reason has become the standard instrument used in most antidepressant RCTs. However, several authors have suggested that the Hamilton is not satisfactorily precise because of the dimensions included, with perhaps too much emphasis on the biological

symptoms (e.g. sleep disturbance, change in appetite, weight gain or loss) (Gibbons *et al.* 1993; Montgomery & Åsberg 1979; Moran & Lambert 1983). It also does not allow for the extremes of those symptoms: i.e. the HAMD only asks about reduced appetite and not increased appetite, for example. A reduced version of the Hamilton scale based on six items, the Hamilton Depression Subscale (HAM-D6), does fulfil criteria for unidimensionality (Licht *et al.* 2004). The HAMD emphasizes the somatic and behavioural or performance aspects of depression, rather than the cognitive symptoms, which are more slowly responsive to drugs than the former (Senra 1995; Lambert *et al.* 1986). In drug trials this could make a drug look better than it actually is if the patient population has more somatic symptomatology. Rather than factor analysis, Montgomery and Åsberg used standard methods of correlation to analyse the changes of scores, interrater reliability and validity. Correlations were found to be reasonably high with values for r being greater than 0.7. The scale has fewer items for somatic complaints thus concentrating on what might be regarded as the core symptoms of depression. Although internal consistency was not specifically tested, Cronbach's alpha has been estimated to be approximately 0.8 (Maier & Philipp 1985). Maier and Philipp also suggest that shorter scales have better homogeneity. When compared with the HAMD, the MADRS is able to differentiate better between responders and non-responders, indicating that it can discriminate changes with greater sensitivity.

Out of the self-rating scales, the Beck Depression Inventory is possibly the most often used in clinical practice, although the Edinburgh scale is utilised in specific situations (post-natal period). The Beck is another scale able to track changes over time and was subjected during its development to several standard statistical tests that showed it has good internal consistency and reliability. The authors considered the usual methods of estimating inter-rater reliability as inappropriate because of the difficulty in deciding how much time should elapse between test and retest (Beck *et al.* 1961; Streiner 1993). They therefore used a modified method that they suggested overcame the problems associated with either a short interval between tests (the subject might remember some answers) or a long one (the subject may be more or less depressed at the second interview). (Although this is precisely what a clinician needs to know in clinical practice.) In the former situation, there would be a higher correlation while it would be lower in the latter. They overcame this by clinically estimating the depth of depression at each interview, finding that there were parallel changes in the scorings. (A similar procedure is seen in RCTs using Clinical Global Impression to evaluate overall severity of illness.) This showed the inventory was mirroring the patient's clinical state. A similar technique was applied to test inter-rater

reliability. The 4-point Depth of Depression rating they used showed an agreement to within one degree of disparity in 97% of cases. However, Beck *et al.* report problems with differentiating the depth of depression of two presentations: one regressed and not eating, the other not regressed but actively suicidal. There was also no validation of the Depth of Depression other than clinical. This could be open to subjective error but the authors attempted to reduce this by comparing 'blinded' results. A pair of doctors would perform a separate assessment for each patient and immediately compare notes for discrepancies.

Two studies were undertaken. In the second study only four psychiatrists were involved in rating patients. They improved their assessment as they gained experience, resulting in greater precision in their judgements. This contrasts with the first study in which there was a fifth psychiatrist assessing patients. This may have skewed results: fewer false positives and false negatives were found in the second study. Hamilton also noted that the precision of judgements by clinicians improved (as indicated by correlations) as they performed more ratings. However in the HAMD, there does not seem to have been any other analysis performed to check for reliability or consistency.

A later analysis of the Beck (Beck *et al.* 1988) showed that the concurrent validity using Pearson product-moment correlations was quite high (mean 0.72) while internal consistency as measured by Cronbach's alpha was also found to be high (mean 0.86). The content validity was good as six items of the BDI reflected nine of the DSM-III well.

The Carroll Rating Scale was designed as a self-rating version of the Hamilton. The scale was modelled on the original Hamilton scale but the Carroll includes two extra values within Item 9 (agitation) from the 1967 revised version of the Hamilton. This revision also includes the four extra items which Hamilton did not consider part of the main scale scoring in 1960. The authors found the correlations within one scale were matched very closely on the other when individual items were compared with total scores. Therefore items on the CRS were reflecting the corresponding ones on the Hamilton with the conclusion that the CRS is an appropriate self-rating version of the Hamilton. The high correlation with the BDI suggested to the authors that the CRS would be a viable alternative, particularly as it also correlated well with the Hamilton. However, 52 items might deter some patients completing the scale properly. This contrasts with the Beck, which has 21 items and, although originally intended as an observer-rated scale, has become a standard self-rating scale as well. The statements of the CRS, in attempting to mirror the HAMD, assess severity of illness. It is also interesting to note that the authors found four items (libido, hypochondriasis, loss of insight and loss of weight) that were so weakly correlated with global severity as to have low predictive utility in either scale. This

begs the question of why the Hamilton has not been revised in the light of such poor utility and might indicate such somatic items are not necessarily distinct components of depression.

The Zung Self-Rating Depression Scale was developed from analysis and summary of statements derived from diagnostic criteria developed by three other authors who had used factor analysis. Zung and Durham found basic similarities in the symptomatology of depression as described by these authors, some of which map quite well onto DSM-IV. Only 56 patients were tested over a 5-week period, 25 of whom were eventually diagnosed as having another disorder. This is a small number of subjects for a valid statistical analysis. However the t-tests showed significance for controls vs depressed untreated subjects whilst there was no significant difference for controls vs depressed treated. Unlike the CRS and Beck, the Zung scale focuses on how much time a symptom has been present, rather than its severity.

The last two scales, the EPNDS and the HADS were designed for particular clinical situations. The Edinburgh Post-Natal Depression Scale was devised to help confirm a suspected diagnosis of post-natal depression in community settings. Validity was determined by comparison with the Research Diagnostic Criteria of Spitzer (Spitzer 1978). No details on statistical analysis are given, except for mention of split-half reliability and Cronbach's alpha. No correlations for the two interviewers involved are presented (although most of the women were interviewed by a single investigator). There is also no clear indication of the correlation between the self-rated and interviewer rated scores. It was noted that the scoring could be influenced by the presence of family members during the interview and also by the subjects' personality, in that the expression of the depression would be different depending on their attitude towards the illness. This could well be relevant when rating with other scales.

For the Hospital Anxiety and Depression Scale, the scale items were based on the psychic symptoms of depression in an attempt to eliminate the somatic symptoms which may well appear in anxiety-provoking situations such as medical out-patients or physical illness. The scale also deliberately set out to differentiate between anxiety and depression in a specific setting in which such differentiation may be important. As the scale does not include the more serious aspects of depression, such as suicidal thought, it may well miss serious mental illness, although a depression score may indicate a patient has a potential diagnosis needing further investigation. The authors were concerned about the relevance of time to the completion of the scale, so they suggested patients consider how they have felt for the previous week to avoid the anxiety and subjective feelings of an outpatient clinic.

They also avoided the middle-response bias when a subject chooses the average response rather than towards the extremes by having only four responses to each statement.

Reliability & Validity of The Scales

Table 3 (page 46) summarises the scales and the diagnostic criteria. Only the Research Diagnostic Criteria (RDC) and the Feighner criteria account for the majority of symptoms of depression. This is not surprising as the RDC was developed from Feighner's earlier work developing a diagnostic system from clinical research experience and validated by follow-up and family studies (Feighner *et al.* 1972; Spitzer *et al.* 1978). Of the remaining two diagnostic systems, DSM-IV has the more comprehensive coverage of symptoms than ICD-10.

From Table 3 it can be seen that the rating scales do not measure the same symptoms or to the same extent. Nor do they all map onto the diagnostic criteria in the same way. It could be argued that a rating scale does not have to map onto the diagnostic criteria precisely as much depends on which symptoms are considered the most indicative of depression. Different weightings within each scale are applied to different diagnostic items or symptom clusters. All scales are designed with a total score as the outcome but it is possible that there may be changes in subscale scores from one measurement to the next, which may give a clinical indication as to how the patient is progressing but actually make no or little difference to the total score (Leon *et al.* 1993), as an improvement in one subscale score may be offset by worsening in another. This gives the impression that the depression of the patient is not lifting when, in clinical and subjective terms, the subject is perhaps feeling somewhat improved. In RCT reports, often only the total score is reported as the primary outcome, and certainly the industry only usually refers to total scores in product literature. A total score assumes that the scale is homogeneous. Analysis of the subscores for a population of patients in a trial may be relevant as a sub-group of those patients may have item scores that differ from another sub-group of the same population, although both groups have the same total score. This may infer that the two groups have the same degree of severity or type of depression. There is also the problem of statistical power if the sub-groups have small numbers: total scores that encompass the whole population have larger numbers that should be more statistically relevant.

Table 5. Statistical Tests used to validate scales

Scale	Reliability		Concurrent Validity	Internal Consistency	
	Test-Retest	Interrater		Cronbach's alpha	Correlation, r
HAMD (1960)	NT	0.9	NT	NT	?
MADRS	NT[1]	0.89	0.7	NT	0.94[2]
Beck	Change of clinical DofD paralleled test score	Indirect % measure	•[3]	NT	0.86 [4]
Carroll	NT	NT	0.80 [2]	NT	0.87 [5]
Zung	NT	NT	NT	NT	NT
EPDS	t-test	NT	Cf with RDC	0.87	0.88[5]
HADS	Different patient sets	NT	NT	NT	0.30-0.60 [6]

[1] Correlation of change/item vs overall change (17 item scale)
[2] Correlation with HAMD
[3] Kruskal-Wallis 1-Way Analysis of Variance, Mann-Whitney U-Test
[4] Correlation between odd and even categories, Pearson r
[5] Split-half reliability
[6] Spearman Correlation of Ranks
NT = not tested

Table 5 summarises the statistical analyses used for assessing the scales' robustness. Reliability was determined by looking for relevant reliability tests including test-retest and inter-rater, and for internal consistency. The first two are usually calculated by using Pearson's correlation coefficient, either between the test and retest scores or between the scores for the two raters. Internal consistency is ideally estimated by Cronbach's alpha, which should have a value of 0.7 or greater. This value is prone to increase with increasing numbers of items in the scale. Only one of the reports used this statistic to validate internal consistency (Cox *et al.* 1987). Spearman's rank correlations were also used.

By inspection of Table 5, it can be seen that the two tests for reliability (test-retest and interrater) have not been employed for all scales. Split-half reliability was used for three of the scales to determine the internal consistency, although Beck also used Kruskal-Wallis Non-Parametric Analysis of Variance by Ranks. In spite of Cronbach's alpha being

a standard test of internal consistency (Streiner 1993), it only appears to have been used in the Edinburgh scale. The complex factor analysis employed by Hamilton was not used by the other investigators, although it has been employed in more recent analyses of some of these scales (Dunbar *et al.* 2000; Enns *et al.* 1998; Galinowski & Lehert 1995; Hammond 1998; Lovibond & Lovibond 1995; Osman *et al.* 1997; Parker *et al.* 2003; Pop *et al.* 1992; Rocca *et al.* 2002).

DISCUSSION and COMMENT

Scales have been constructed on the basis of their authors' own perspective of depression: cognitive, somatic and behaviours all being assigned varying degrees of emphasis within a given scale (Snaith 1993). The weighting given to these will alter the sensitivity to change and the rate at which the scales respond to change, particularly relevant in trials where small differences between two drugs are being detected (Montgomery & Åsberg 1979; Senra 1995). Table 3 (page 46) describes the weighting of the scales as percentages and shows that the scales put different emphases on the various symptoms associated with depressive illness. But what should be detected for the purpose of clinical trials as opposed to the outcomes sought after in clinical practice? Lifting of mood is the prime target of treatment, whether for a clinical trial or for clinical practice, but there are also changes in symptoms (e.g. weight loss, sleep) that are not directly related to the core items, albeit important to the overall relief to the patient, and perhaps changes in attitude to illness. These other, non-core, symptoms are possibly not so important for assessing the efficacy of new antidepressants in clinical trials. However, it may perhaps depend on the desired outcome of drug treatment and whose outcome: patient, doctor, drug company, or healthcare purchaser. The choice of scale would be important to a company: one which demonstrated a change in non-cognitive symptoms (e.g. Hamilton) before changes in the core, cognitive, symptoms might well be preferable in a randomised controlled trial to one that demonstrates the later changes in the cognitive symptoms (as would be the case with the MADRS). Having a wider range of symptoms may also lead to reporting of the non-efficacy symptoms, which may confuse or obfuscate the true picture.

Out of all the items that might be distinguished from a sample population, only the more relevant should be used to construct a scale otherwise it would be difficult to administer and time-consuming if it was to be used in normal clinical practice. Scales must also be easy to administer as they may possibly be used by untrained raters. For self-rating scales, a long series of statements will lead to patients being unwilling to complete the

scale properly. So a scale must only include those symptoms that describe the illness, and must be accurate, reliable, consistent across a given population and able to distinguish true depressives. For example, patients with depressive illness will have similar symptoms to those with medical illness who also have low, but understandable, mood with somatic symptoms indicative of depression but not actually clinically depressed (Zigmond & Snaith 1983). A scale must also be able to differentiate between the low mood of 'true' depression from the low mood that is due to a low threshold of the individual to cope with difficult circumstances. Personality and attitude inventories may have an important role to perform in these situations, particularly in clinical trials where it is important to include patients with clinical depression. Individuals with low mood that is not a result of clinical depression may perhaps benefit from non-pharmacological intervention. If RCT exclusion criteria do not filter out patients with a personality disorder then there is the risk of investigators including 'less' organically depressed patients who may respond to antidepressant treatment differently. They may well also respond to the more intense attention from all the trials visits.

In RCTs, patients are usually diagnosed on a clinical basis and classified by a diagnostic system such as DSM-III (and its revision), DSM-IV, ICD-10. These systems only describe the symptomatology of the illness and label it, and do not give a scoring of severity or intensity of illness. Some trials have also used the Research Diagnostic Criteria (Spitzer *et al.* 1978) (instead of, or as well as, DSM, CGI). If depressed according to these criteria, the patient is then rated on one or more scales. Perhaps the ideal rating scale would map itself onto DSM in such a way that the criteria for making the diagnosis would match those for rating the depth of depression. However, no one rating scale appears to do this (Table 3), although the HAMD, CRS and Zung come close. This may mean that the diagnostic criteria are also multidimensional and not specific for the core symptoms.

It is also usual in clinical trials to further categorise the overall severity of depression using the Clinical Global Impression (CGI) (Guy 1976). This gives an overall impression of the severity or intensity of the depressive episode but could be very subjective and prone to inconsistent rating behaviours. The severity of illness scale is poorly constructed with insufficient verbal labels and logical relations (Beneke & Rasmus 1992). Beneke and Rasmus have also described the efficacy index of the CGI as 'misleading' and 'redundant'. It is notable that few, if any, RCTs use this part of the CGI. It could be argued that improvement is being measured by the use of rating scales so that this global measurement is perhaps superfluous. It would be expected that the CGI score for global improvement would get better as the depression rating improved. However, it

helps to anchor symptoms at the time of assessment: Beck *et al.* used the same technique when developing their scale (Beck *et al.* 1961). What might be more useful is a measurement of functioning; i.e. quality of life (QoL). This could be useful in financial decision-making, as taking into account the possibly increased ability of patients to return to work due to better QoL, would potentially reduce the financial burden on the health economy.

This review has considered two types of scales: expert- and self-rated, although self-rated scales are little used in clinical trials. There are problems with both. With expert-rated scales, when more than one rater is involved, there is always the possibility of differences in scoring between physicians. This may be due to unfamiliarity and/or lack of experience of using the scales, thereby yielding different scores, particularly if different centres are used as happens in clinical trials. There may also be a difference in perception of severity between clinicians; the severity of an illness may be overestimated by inexperienced doctors while more experienced psychiatrists might describe it as moderate. Using centres in different countries introduces the further possibility of cultural and attitudinal differences in describing and interpreting depression (Senra 1995), although some studies have indicated this may not be so (Åsberg *et al.* 1973; Ramos-Brieva 1988; Wickberg & Hwang 1996). There is also the possibility of differences in meaning of items when scales are translated from English into another language. However, this does not appear to be a problem (Pop *et al.* 1992; Åsberg *et al.* 1973). Content differences between self-rated and expert-rated scales may alter the sensitivity of treatment interpretation as it has been shown that effect-sizes can be quite different depending on which type of scale is chosen, although the results can be contradictory (Lambert *et al.* 1986; 1988).

Self-rating scales suffer from the possible exaggeration or minimisation of symptom severity by the patient in some circumstances (Cox *et al.* 1987). Subjects may consider themselves worse than a doctor would rate them. Conversely, some patients may consider they have a large improvement when compared to physician rated scoring (Lambert *et al.* 1988). This may be due to the way patients report their symptoms to the doctor. A patient's description of their symptoms may not necessarily be borne out on a self-rating scale and, if the patient is completing a scale, on how they therefore interpret scale statements. The interpretation by the doctor of the patient's affect and assessment of the description of how the patient feels may be different. Also, the change in scores between ratings is not as great. For example, the BDI showed greater improvement in subjects' depression than did the HAMD or Zung (Moran & Lambert 1983), although the authors used a box-score method that does not account for the different amounts of change

between different scales (Lambert *et al.* 1986). It may be better to use effect sizes that are statistically based indices of the magnitude of effect or change (Lambert *et al.* 1988; Senra 1995; Faries *et al.* 2000).

It is salutary to note that patients with quite different mental illnesses can have the same score on a rating scale. Cooper & Fairburn (1986) compared the high scores from bulimia nervosa patients with those of patients suffering from primary depressive disorder using the MADRS. Such scores inferred the bulimics were depressed. Content analysis of the scores revealed the bulimics scored highly on the 'tension' and 'pessimistic thoughts' items, while the depressed patients had high ratings for 'observed depression', 'reduced sleep' and 'suicidal thought' items. Again this raises the issue of which symptoms should be included in a 'definitive' rating scale for depression. Rather than considering the somatic features too much, perhaps a focus on the cognitive symptoms would be more appropriate (Gibbons *et al.* 1993). This would make the scale more unidimensional and less prone to influence from somatic symptoms which can occur in other, non-depressive states. The exclusion criteria used in clinical trials should ensure a homogeneous population with a similar depth of depression and similar symptomatolgy, so this problem should not, in theory, arise in randomised controlled trials. However, this does not account for subjects with personality traits that might predispose them to low mood by having a pessimistic view on life. Such traits would not be amenable to pharmacotherapy and therefore would potentially skew results in trials.

Scales may be assessing different aspects of depression. The HAMD may have content differences from other scales, e.g. the BDI or Zung. HAMD does well with somatic symptoms but not so well with cognitive and affective changes as does the BDI (reflecting Beck's cognitive view of depression). It is too multidimensional. These symptoms resolve at different times, so that the BDI will detect changes shortly after the HAMD will have picked up on the somatic symptom change. The HAMD may also be showing its age. Since it was devised in 1960, social and temperamental attitudes have changed, so the scale questions may need updating to reflect this. Careful selection of scales is required, as it appears there are differences between the scales and what they measure (Snaith 1993). It is possible that a patient may be misdiagnosed with depression using one scale when another may be more appropriate (Professor Oyebode: personal communication).

There is a gender difference in the incidence of depression and perhaps this should be taken into account when conducting clinical trials The incidence of depression in females is some two times higher than that in males but is that depression different in some way to that found in males and if there is a qualitative difference could this have an effect

on RCT outcomes? This may have a bearing on the choice of scale in a population (Areias *et al.* 1996).

Why do most investigators use HAMD while others use MADRS despite various other scales being developed over the years? The HAMD & MADRS are the most commonly used depression rating scales used in RCTs with the Hamilton the more commonly used of the two. Out of forty-one trials in one review by Furukawa *et al*, twenty-one used the Hamilton, only six used MADRS, while the rest used other scales (Furukawa *et al.* 2002). The reasons for choosing a particular scale in a trial are never given. Furthermore, there are variations on the HAMD used by different investigators, developed as seen appropriate by the researchers. In many trials, the 'updated' or otherwise altered version is not specified in detail so that it is impossible to say how different the 'new' scale is from the original. There are instances when the HAMD scale used is different in some way but the reference is for the wrong version. For example, Hamilton's 1967 version (21 items) is referenced but the scale apparently used is the 1960 original (17 items). Hamilton did not intend to use the last four items as part of the overall score. In these circumstances it makes comparison of trials more difficult to interpret (Grundy *et al.* 1994).

Some trials use more than one scale, often HAMD and MADRS. This may be advantageous, as the instrument selected may influence the outcome due to the different focus of each scale, which would make the comparison of conclusions with other studies using different scales difficult if not impossible. Careful selection of scales is required, as it appears there are differences between the scales and what they measure (Snaith 1993). There seems to be no clear reason why the HAMD is used so widely, except that it does have consistently high inter-rater reliability (Ziegler *et al.* 1978; Moran & Lambert 1983). The MADRS also has good inter-rater reliability (Davidson *et al.* 1986) and has been shown to be more sensitive to change during treatment. Trials in recent years have tended to use MADRS as the primary outcome scale.

What is needed, perhaps, is a twin-score system. One score would measure the core symptoms of depression (i.e. depressed mood, anhedonia, loss of energy) and a second to measure the somatic symptoms which, it might be argued, arise from the underlying low mood (e.g. sleep disturbance, loss of appetite, loss of weight). The former score would reflect the core symptoms found in diagnostic classifications. The second score would add to the first by reflecting the impact on daily living and general well-being. A measure of personality or attitude should also perhaps be included in addition to the secondary symptom score, as these can affect an individual's ability to cope with serious

illness and would perhaps give a measure of the subject's tendency to over-estimate severity of symptoms. There may be a connection between the early onset of major depression and personality pathology (Ramklint *et al.* 2003).

The problem is to try to define what the core symptoms are and what the main outcome should be. A key criterion for diagnosing and measuring depression would appear to be a subjective and objective lowered mood, although it is not a mandatory requirement of either DSM-IV or ICD-10. Patients will describe many symptoms, both physical and psychological, as part of their depression but these descriptions are not necessarily helpful or relevant in determining the core symptoms or level of severity. Nor are these symptoms necessarily indicative of depression. In terms of outcome, trials usually only require response and not remission of symptoms to demonstrate a drug's efficacy, while in the real world, remission is the important goal.

We certainly need to be more aware of the utility of the rating scales used in clinical trials of antidepressants and the possible problems associated with them. In spite of its age, the HAMD is still a popular choice for RCTs. There seems to be no clear reason for this, except that several factor analyses have vindicated its ability to tap the symptoms of depression. Similarly for the MADRS, which has shown its ability to give reliable ratings in RCTs. The scale was chosen as the primary rating scale for the escitalopram RCTs.

With the impending updates of both DSM and ICD, and given that there is an attempt to form a consensus between them, it perhaps time that the rating scales were also updated to reflect a changing population and social structure. In addition, it would appropriate to consider what items should be included in a new scale, perhaps excluding those symptoms that are not specific to depression.

So, before deciding if the outcomes of a paper or set of papers are appropriate for a clinical situation, we need to decide if those outcomes reported are accurate and comparable. This is particularly important when interpreting the results of a meta-analysis if the papers used for it contain slightly different outcomes depending on the version of the HAMD used, the use of different rating scales in trials being compared with each other, or the use of secondary outcomes instead of primary ones in report conclusions.

CHAPTER 4

ECONOMICS OF DEPRESSION AND ANTIDEPRESSANTS

Introduction

Healthcare costs have to be controlled and the principles of economics can be used to determine the allocation of scarce resources between competing needs. The costs and economic consequences can be estimated using several pharmacoeconomic analytical methods. Costs associated with acquisition of pharmaceuticals, hospitalisation and professionals' time are relatively straightforward to estimate. However, the difficulty arises when trying to apply a financial cost to intangible outcomes: health gain, disease burden, quality of life. The illness will have an impact on daily living, work and productivity, but these aspects of life are difficult to measure as they are very subjective. However, there is a need to analyse the financial and quality of life issues in health economic terms (usually costs) to estimate that impact so that resources can be appropriately allocated.

There are three main analyses used in health-economics, which differ in the way that health outcomes are assessed and measured (Table 6). Cost of illness (COI) and cost-minimisation are not recognised as economic evaluations as all factors are allocated costs and there is no evaluation of the outcome of intervention. Which factors are included in COI will depend on the perspective of the analyst.

Table 6: Pharmacoeconomic analyses

Method	Outcome Units
Cost-effectiveness	Natural units
Cost-benefit	Money
Cost-utility	QALY or DALY

QALY = Quality-adjusted life years; DALY = Disability-adjusted life years

The first part of this chapter reviews the types of economic analysis and modelling techniques that can be used, before presenting a systematic review of studies that have analysed the cost and economics of depression that might be used to inform healthcare commissioners in deciding resource allocation. There then follows an analysis of the pharmacoeconomic studies relating to treatment with antidepressants, with particular review of the drug chosen as an example for this study, escitalopram. This drug was chosen because it was brought to market with a minimum dataset and little economic data, although this was subsequently followed up with a number of pharmacoeconomic studies, which will be discussed.

Economic Models

Cost of Illness

Cost of illness (COI) has been used to quantify the direct and indirect costs resulting from an illness by using, most commonly, the prevalence method, which estimates the total annual cost of all individuals with the disease. Prevalence is defined as the proportion of the population affected by depression at a given point in time (point prevalence) or period of time (period prevalence). A second method for estimating the cost over a period of time is the incidence approach, which only considers the costs associated with newly diagnosed individuals during that time. Generally, these costs have been the direct costs of treatment (eg surgery, physiotherapy, use of emergency services) and pharmaceutical costs, and the indirect costs from lost productivity due to absenteeism or presenteeism (the reduction in productive capacity while at work), increased morbidity, and increased benefit payments (Stoudemire *et al.* 1986). Some researchers have included the reduction in a patient's productive capacity while at work during depressive episodes (Greenberg *et al.* 1993; Beuzen *et al.* 1993). COI studies concentrate purely on the expenditure involved in treating an illness and take no account of the outcomes of treatment. The perspective for this type of study is often that of the healthcare purchaser but can estimate the impact of a disease on the broader society. As it does not take into account the patient outcome (in the present discussion, relief of depression or the associated sequelae of unsuccessful treatment), any costs associated with a good outcome are not accounted for: the perspectives of neither patient nor provider are acknowledged.

Cost-Minimisation Analysis

Cost-minimisation analysis compares two treatments in cost terms only because their outcomes (effectiveness and safety) are identical. It therefore becomes a basic comparison of drug acquisition costs. Two issues arise. First, care must be taken with potentially subjective outcomes such as rating scales whose results need careful interpretation. (Measurements in physical medicine, such as biochemical parameters that have clear results or blood pressure, which has defined ranges of measurement, are less subjective.) It is important to ensure that the results from different trials relate to the same outcome. Second, care must be taken when collecting data sets, as the inclusion of a new compound with a significantly better outcome or adverse event profile could be construed as being sufficiently different from comparators to prevent its inclusion in a cost-minimisation analysis, as the analysis should be comparing essentially identical products which are to be separated by cost: cost-effectiveness analysis would be more appropriate.

Antidepressants recently marketed have little (if any) greater efficacy than more established ones (Anderson & Tomensen 1994). However, it could be argued that side-effect profiles may differ enough for cost-minimisation to be inappropriate, although NICE appear to have taken this approach in Clinical Guidance 023 (NICE 2004).

Cost-Benefit Analysis

Cost-benefit analysis is derived from economic theory and compares the incremental cost of using a health care intervention (antidepressants, for example) with the benefits of using that intervention compared with an alternative or no intervention. Both net costs and the benefits are expressed in monetary terms. Benefits are often valued by using willingness to pay, which may depend on the ability or acceptability to pay for an intervention. The net cost of the intervention includes all the direct and indirect costs less the similar costs for the alternative intervention. The analysis end-point is the benefit minus the net cost, i.e. the net benefit. A positive net benefit usually means the intervention should be funded. The cost-benefit ratio (ratio of the net cost value and the benefit value) that is sometimes calculated in analyses is not recommended for use as a decision criterion (Berger *et al.* 2003).

This analysis generally takes a societal perspective and tries to include all costs, but calculating indirect costs can be difficult and sometimes controversial. There may also be ethical concerns and difficulties about using monetary values on life and health state.

However, CBA does have two advantages. Like cost-utility analysis, CBA can compare two interventions that have different outcomes: the decision rule employed is to opt for the treatment with the higher benefit. Secondly, CBA is the only method with a single decision rule for evaluating single interventions: funding should be found for a positive net benefit.

Cost-effectiveness Analysis

CBA differs from cost-effectiveness analysis because in the latter, two or more treatments are systematically compared using the cost and outcomes of each. This analysis usually has a narrow perspective, for example that of the purchaser. Outcomes across the interventions to be compared are measured in units that are related to the clinical outcome, such as symptom-free days gained, life years gained, time to remission of depressive symptoms. When comparing two treatments (for example, two antidepressants, or antidepressant treatment versus psychotherapy), the incremental cost-effectiveness ratio (ICER), which is a measure of the additional cost per unit of health gain, can be calculated:

the difference in cost (incremental cost) of each is divided by their difference in outcomes (incremental effect). When comparing more than two treatments, systematic pair-wise analysis of the ICERs is used after eliminating the ones obviously dominated due to being more costly and less effective.

Dominance arises in a cost effective analysis when one strategy is more effective and/or costs less than alternative ones. The alternatives are ruled out and are said to be 'dominated'. In simple dominance, an alternative is both more effective and less costly, while if there is a more effective but more costly alternative, which provides better value for money, there is said to be 'extended dominance'. The terms that remain after all dominated terms are eliminated form the 'efficient frontier' as they are all potential technologies or programmes that could be used. The decision as to which should be chosen is based on the threshold cost per QALY of the decision maker. Simple dominance is relatively easy to apply and is not controversial in its application. Extended dominance is more complex to apply, as the alternatives will have complex budgetary implications.

Cost-Utility Analysis

Probably the best methodology for analysing costs and benefits is cost-utility analysis, which not only estimates costs but also accounts for outcomes in terms of gain in life years and health utility or preference. This combination is usually expressed as the cost per quality-adjusted life year (QALY). Using the QALY as a common denominator across studies could allow cost-utility measures to be compared across studies, and an acceptable 'threshold level' of cost/QALY set (ISPOR Book of Terms, p45). The previous methodologies (CBA and CEA) discussed above are prone to biases; cost-benefit analysis can lead to inequalities between differing groups of individuals when the human capital approach is used as this measure depends on the ability to earn, while the intermediate outcomes that might be used as the denominator in cost-effectiveness analysis across different disease states cannot be compared with each other.

Quality-adjusted Life Years

Use of the cost/QALY is not always straightforward. Cut-off threshold values used to determine cost-effectiveness may differ for different situations; e.g. when comparing chronic against acute conditions. Otherwise, the incremental cost-utility of, say, a new antidepressant (An) vs an older one (Ao) can be calculated from

$$(\text{Cost(An)}-\text{Cost(Ao)})/(\text{QALY(An)}-\text{QALY(Ao)})$$

where

QALY = number years survival x utility value

There is usually some trade-off between the quantity of life (years survived; mortality) and the quality of life (morbidity), such that high quality may only be for a few years or the individual might live longer but with a poorer standard of living. A QALY is adjusted by a preference-based quality weight, which is usually determined from a utility scale that measures the preference of health state. The population preference may vary but is currently considered to be that of the community. This scale ranges from full health (1.00, unity) to death (zero, 0.00). Negative scores indicate a state worse than death, where quality of life is so poor the individual sees death as preferable. The number of QALYs gained is usually not equal to the number of life-years gained. Time trade-off may be used to estimate utilities. This method asks an individual to express their preference for a particular outcome of health state in terms of the maximum loss, expressed as reduction in healthy life expectancy, they would accept in order to avoid the loss. It is also usual to apply discounting to the calculation, to adjust future costs and benefits to the current market value. By combining it with cost data, the marginal cost per QALY can be estimated. The objective of a CUA is to estimate the marginal cost for a given QALY being delivered by an intervention.

Methods for quality of life measurements in psychiatry were in a primitive state in the late 1980's (Wilkinson *et al.* 1990). The measurements at that time were considered incapable of accurately characterising psychiatric problems reliably using a two-dimensional index, such as the descriptive scale used by Wilkinson *et al.*, which only

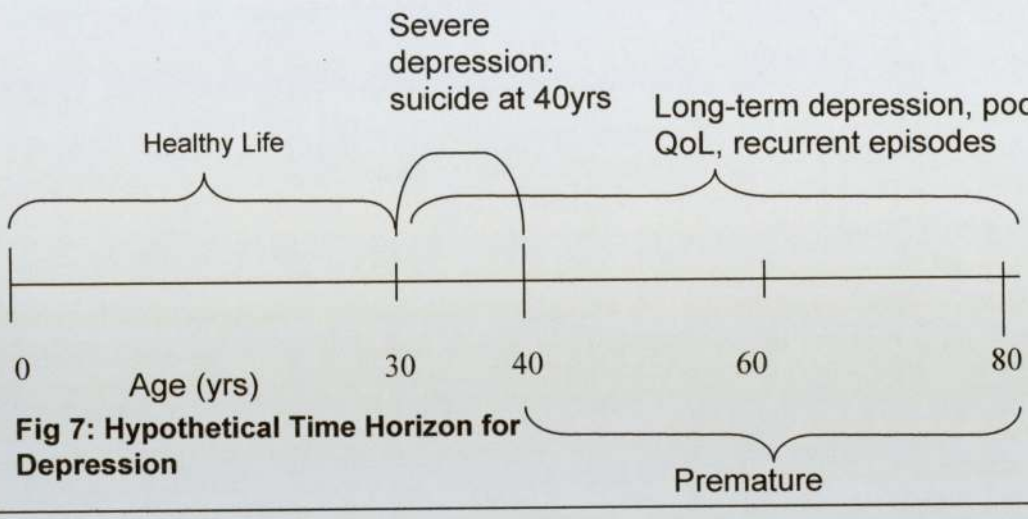


Fig 7: Hypothetical Time Horizon for Depression

accounted for the levels of distress or disability. The measures they considered only looked at clinical outcomes, whereas psychiatrists usually regard social outcome as an important indicator as well. A range of specific QALYs in psychiatric care will be needed to assist in rational resource allocation. The situation had apparently not improved by the early part of the new millennium, as a review of a large database held by the Harvard Centre for Risk Analysis revealed a lack of cost-utility research in depression (Pirraglia *et al.* 2004). A more recent paper by Mann *et al.* in 2009 suggests that two generic preference measures for quality of life, the EQ-5D and SF-6D, are able to reflect changes in health state that mirror the improvement in depressive symptoms (Mann *et al*, 2009).

The World Health Organisation now promotes the use of QALYs as the standard approach, although the version it uses is called the DALY (disability adjusted life year), particularly for long-term health outcomes. The DALY was developed by the World Bank in 1993 and further refined by the WHO to estimate the global disease burden across different diseases. Some care is needed in assessing QALYs, as there are several quality of life instruments which can be used (eg EuroQoL EQ-5D, Health Utilities Index), the results from which cannot be directly translated one to another, as the quality weights used for assessing utility in the tools are measured in different ways. For example, the two most often used utility measurements estimate life outcomes, in terms of maximum loss to an

individual, either as the risk of a particular bad outcome (standard gamble) or as the reduction in healthy life expectancy (time trade off).

Essentially, QALYs and DALYs are similar in that they estimate the impact of disease in terms of mortality and morbidity. However, the DALY considers premature death for its mortality measure rather than the QALY measure of death, which is the eventual death of an individual. The DALY also uses an external standard life expectancy (based on data from the country with the highest life expectancy, Japan) from which premature death is calculated. In Figure 7, depicting possible scenarios in depression, a hypothetical patient has a healthy life until thirty years of age when they suffer their first episode of major depression, Looking at a worst-case scenario, this individual never fully recovers, ultimately committing suicide ten years later. This would be expressed as years of life lost (YLL):

YLL = average life expectancy – age at death.

From Fig 7, this would be 80-40: i.e. 40 years of life lost.

A QALY calculates the life years by using the number of years in a given health state multiplied by a quality weighting score. For the individual in the example, this might mean 30 years in full health, followed by 10 years in a lower health state with a value of 0.3, followed by death:

$(30 \times 1) + (10 \times 0.3) = 33$ QALYs.

If successful treatment brought about a higher health state value for 5 years (say, 0.8), followed by a severe relapse to give a health state of only 0.3, then death after 5 years, the QALY $= (30 \times 1) + (5 \times 0.8) + (5 \times 0.3) = 35.5$. This crude example shows that the patient only has a small improvement in life years gained.

DALYs also account for the quality weighting for morbidity by estimating the years lived with (or lost to) disability (YLD). This takes into account disability weights (derived from various non-fatal conditions), age weights (the importance of healthy life at different ages), and an estimation of the value of health gains in the present compared to the value of future health gains. Adding YLL and YLD gives the DALY:

DALY = YLL + YLD.

The DALY approach has been criticized for not being representative of the societal perspective, since the preference weights were based on person trade-off scores from an expert panel rather than those of society (Health Care Cost, Quality & Outcomes. ISPOR 2003). QALYs are usually derived from disability weights that are society preference-

based utilities as in the EQ-5D. The weighting of health states according to age, where lower weights are attributed to the young and the old, may introduce bias to these populations (Anand & Hansen 1997), although in the UK, the Citizens' Council has said that NICE should not let any of its decisions be influenced by the age that patients might be when a particular intervention is being considered (Citizens' Council Report, 2004). Further controversy concerns the weighting in favour of the population which has no disabilities, inferring that the life years of disabled people are worth less than those with no disability (Arnesen & Nord 1999). Disability could include mental illness. A similar observation was made with respect to QALYs in 1990. QALY-based judgements may bias against individuals with poorer, long-term outcomes, such as mental illness (Wilkinson *et al.* 1990). The outcomes in mental illness are less well-defined and do not necessarily reflect a positive end-point of complete well-being: such illnesses are often life-long with periods of being relatively well interspersed with periods of ill-health and the associated poor quality of life. The generic QALY was therefore considered to be inappropriate for aiding decisions on healthcare resource allocation and more specific measures should be constructed (Chisolm *et al.* 1997). However, use of the EQ-5D and SF-6D health related quality of life measures appear to be sensitive to changes to changes in health states, at least in depression (Mann *et al.* 2009).

Productivity Costs

When estimating the burden of an illness on society, one of the indirect costs that should be accounted for is the cost of lost production or productivity. This is difficult to estimate as a price has to be placed on the reduction in output from an individual who is either at work but working at a reduced capacity or is off sick from work and therefore not contributing. The latter is absenteeism, the commonly considered situation when a person not at work is not contributing to output. When an individual is working but at reduced ability has been termed presenteeism.

There are two approaches to help understand this cost. The human capital approach estimates the productivity cost in the absence of market prices. The value of human capital is estimated as the present value of an individual's future earnings. It is used to estimate the indirect cost of illness. Human capital consists of those attributes that contribute to their ability to produce: e.g. knowledge, skills, and health. However, it accounts neither for non-earnings production (e.g. housewives), nor for leisure time. So it may overestimate the value of foregone production. There are other problems with this method, such as wage discrimination for different people doing the same job for different wages.

An alternative approach is to use friction cost. This estimates costs in the friction period, when the person is off sick before it is necessary to replace them. It underestimates the cost of lost production and assumes the worker will be replaced by an unemployed person. However, some experts contend that the human capital approach can overestimate indirect costs. A fundamental concept of friction cost is that the sick worker will be replaced by another one, if the illness is long-term. The friction period is the time from the start of the absence to the time at which productivity is restored to its level prior to the absence (Birnbaum 2005). The actual losses made during this time are those under consideration, not the wages of the individual (as in the human capital approach), which are used as a proxy measure of work output.

Cost Models

Cost models are an alternative approach for determining the potential costs associated with treatment alternatives. A decision tree is constructed and outcomes of given decisions for treatment are analysed by calculating the probabilities of events and their outcomes through the branches of the decision tree (Kind and Sorensen 1995). Decision analysis models are flexible, able to incorporate differing scenarios over different time durations. There are four main steps to analysis: identify and define the decision problem; describe the structure of the problem over time; identify the information needed; choose a course of action (Jönsson & Bebbington 1994). Sensitivity analysis is an essential part of the analytical process to determine the stability of the model. The method does assume that each outcome can be defined as a discrete event with well-defined probabilities. It cannot accurately predict outcomes in non-binary systems; i.e. those which vary continuously over time or between individuals. Guidelines regarding the use of decision analysis modelling have recently been published by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), which help to define more robust modelling (Weinstein *et al.* 2003). A good model will reveal the logical connection between its inputs of data and assumptions, and its outputs of valued consequences and costs. A sensitivity analysis will demonstrate if there are any effects from alternative data and assumptions. To avoid favouring the investigational drug, assumptions are often made, which introduce bias against it. Examples include the use of conservative estimates of efficacy or use of the lowest cost or lowest dose for the comparator drugs. However, not all studies are explicit about possible biases in the model.

Decision analytic or Markov models use various settings for analysis: primary or secondary care, the stage of the depressive episode (acute or maintenance), the type of

treatment (maintenance or episodic). There is usually no differentiation between inpatients and outpatients. The type of depression is usually specified but it can be a sub-set of major depressive disorder.

The criteria used to construct the model should be stated in analyses to enable their quality to be assessed. The sources of data, such as well-conducted randomised clinical trials, epidemiological data, and expert opinion from Delphi panels need to be specified when reporting a cost model analysis. The exclusion of any data sources when estimating parameters should be justified. Where expert opinion is used, it is good practice to demonstrate that these parameters would not unduly influence the outcome: a sensitivity analysis would help elicit this. ISPOR recommend that a sensitivity analysis should have a 'clear statement that results are conditional upon this (these) subjective estimates(s)' (Weinstein *et al.* 2003). Similarly, outcome measures should be clearly defined. The decision analytic model therefore attempts to distil the best data available from observational studies, RCTs, claims databases, case registries, public health statistics and preference surveys, and use that data to simulate what might happen under various decision conditions and outcomes. Retrospective clinical data can limit the generalisability of results to clinical practice, however (Wilde & Whittington 1995), and must be complete and readily available (Frank *et al.* 2001). Costs might include those for drug, hospitalisation, GP visit, and other health professionals.

Expert ('Delphi') panels of general practitioners or psychiatrists may be used to define parameters that can be used to form probabilities for an event in the decision tree, and advise on practice patterns, clinical pathways and treatment strategies, although these may be a source of bias (Frank *et al.* 2001). These estimates rarely examine factors involved with real world outcomes such as variations in treatment delivery, non-compliance, and co-morbid illness (Simon *et al.* 1995). The outcome results of placebo trial patients may be applied to the outcomes of 'untreated' or poorly treated real-world patients, but this ignores the influence of compliance in the latter group. Discontinuation is more likely due to side effects or lack of efficacy.

Economics of Depression Analysis

Method

A search for economic studies for antidepressants and for depression was made using MEDLINE, PubMed and ISI World of Science databases, ranging from 1980 to 2005, using the terms antidepressant, depression, cost-benefit analysis, cost-effectiveness

analysis, cost-utility analysis, cost of illness, decision analysis, decision analytic model, fluoxetine, citalopram, escitalopram, sertraline, paroxetine, venlafaxine, mirtazapine, tricyclic antidepressants. These search terms were felt appropriate to find economics studies for the relevant, commonly-used, antidepressants, and for the economic analyses studying the cost of depressive illness. Further papers and reports about economic studies and terminology were identified from reference lists.

Results

Table 7: Economic Burden of Depression Studies

Author	Stoudemire *et al.*	Greenberg *et al.*	Kind & Sorensen	Jonsson & Bebbington	Thomas & Morris
Publication year	1986	1993	1993	1994	2003
Setting	USA	USA	UK	UK	UK
Year of Estimate	1980	1990	1990-91	1990	2000
Total Cost	$16,300,000,000	$43,700,000,000	£3,389,658,000	x	£9,055,274,000
Direct Treatment Cost	$2,113,325,528	$12,411,650,844	£416,658,000	£222,000,000	£369,865,000
Drug Costs	$138,378,780	$1,175,000,000	£47,280,000	£47,722,000	£310,378,000
DrugCost as % Total Cost	6.5	9.5	1.39	21.5*	3.43
Outpatient Care	$657,919,939	$2,792,057,874	£9,146,000	x	£22,133,000
Inpatient Care	$1,269,471,579	$8,344,973,520	£177,365,000	x	£28,660,000
Mortality Costs	$4,200,000,000	$7,520,869,093	x	x	£562,151,000
Morbidity Costs	$10,028,000,000	$23,800,000,000	£2,973,000	x	£8,123,258,000
GP Costs	$175,716,580	x	£126,399,000	x	£8,217,000
Social Costs	$16,300,000,000	x	£40,183,000	x	x
Working Days Lost	155,937,211	289,425,044	155,177,000	x	109,700,000
Loss of Earnings	x	$7,520,869,093	£2,973,000,000	x	>£8 billion
Loss of Future Earnings[1]	x	$7,520,869,093	x	x	£562,000,000
Deaths (suicide)	16,111	18,446.00	2239	x	2,507
Prevalence Rate: male[2]	2.2	x	14.8	x	40.3
Prevalence Rate: female[2]	5.7	x	53.7	x	102.7
Age Highest Prevalence	25-44	30-34	45-64	x	35-44
No. Cases	4,757,779	x	84,633	x	2,600,000
% Female	72	71	75.5	x	72
Age Highest Suicide Rate	20-29	25-34	x	x	25-34
Acute Hosp Admissions	565,532	498,000.00	24,791	x	x
Psych Hosp Admissions		$104,135	59,842	x	x
Life Years Lost	377,768	x	90,189	x	x

[1] Due to premature death

[2] annual, cases/1000. Except Stoudemire which are 6 month prevalence estimates.

x = not stated

*% Treatment Cost

Economics of Depression

This study found that five key economic analyses have been published since 1980 that estimate the costs of depressive illness (Table 7). Although there are variations in the estimates of burden of depression in these cost of illness studies, there is a pattern indicative of the high costs of the disease to society.

The earliest example of cost analysis in depression was conducted by Stoudemire *et al.* (1986), who developed an analysis based on the Epidemiological Catchment Area Project of 1980 in the USA. Data from this programme were used to estimate the direct treatment costs and the indirect costs due to the lost productivity arising from the morbidity and mortality of major depression. Unlike the other analyses (Table 7) with time horizons for prevalence estimates of twelve months, the time horizon used for prevalence by Stoudemire was six months. Years lost were accounted for, both in terms of activity and of life years but the costs were underestimated, as the indirect costs did not include intangibles such as pain and suffering of the individual and/or their family and friends, and lost future earnings when assessing the cost of mortality. However, the estimates for the US indicate the potential savings to be made if timely and appropriate intervention is made. At 1980 prices, the costs were estimated at $2 billion for total direct costs, $10 billion for lost productivity, and $4 billion for total mortality costs due to lost productivity. Over 75% total costs of depression to society are indirect due to lost social and economic productivity. Drug costs accounted for 6.5% of the total costs.

Taking the conceptual framework of the Stoudemire analysis a stage further by estimating the cost of all types of depression (including bipolar depression and dysthmia, not just major depression) and by also considering the costs of reduced productive capacity while at work (Stoudemire only included costs of absenteeism), Greenberg *et al.* (1993) concluded the total annual cost of mood disorders in the United States was around $44 billion. This is more than twice the estimate of Stoudemire *et al.* but the analysis does include extra categories of depression. Of the approximately $1.2 billion drug costs estimated by Greenberg *et al.* for the treatment of depression in 1990, antidepressants accounted for approximately $890 million. This increased proportion of drug costs may in part be accounted for from the increased awareness of depression resulting from national promotional campaigns. Other costs were $190 million for anxiolytics, $25 million for antipsychotics, and $70 million on other pharmaceutical therapies. The drug costs in the Greenberg study represented 9.5% of the total costs. Compared to the UK estimates for suicide, the Greenberg estimate is high, although it includes all affective disorders. The authors used the human capital approach to develop prevalence-based estimates for direct

costs, mortality costs, and morbidity costs. The original 1980 data used by Stoudemire was updated by Greenberg *et al.* to estimate 1990 values.

Similar estimates for England and Wales based on then 10-year-old data have been calculated (Kind and Sorenson 1993). The authors estimated a 12-month prevalence rate of 35.2/1000, with rates being higher for women, particularly in the 45-64 age group. They estimated general practitioner psychiatric consultation costs to be higher than medical consultation costs because psychiatric consultations take longer. The estimate derived by Kind and Sorenson was £126m, and at that time 40m prescriptions were issued by GPs each year, and dispensed, at an average cost of almost £6 per prescription (Kind & Sorenson). With the total direct costs to the NHS for both primary and secondary care over £416 million and the loss of 155 million working days equating to nearly £3 billion in indirect costs, the total burden due to depression was estimated at approximately £3.5 billion, of which direct treatment costs account for approximately 25%. Drug costs were estimated to be more than £47 million, although this only represents 1.9% of the total costs. It is not clear if these costs were for antidepressants alone or if it included other treatments. The authors' estimates were based on historical data, and prospective data collection would give a more accurate estimation of the scale of the problem.

In contrast to the Kind and Sorensen approach, Jonsson and Bebbington (1994) reported only the direct costs for cost of illness, a figure of £222 million, derived from unit costs and total number of events in 1990 (Jonsson and Bebbington 1993), together with a total drugs bill of approximately £41.7 million. This proportion of the cost of illness due to drug costs accounts for 21.5%. They considered that the data available for calculating indirect costs were unreliable for the purpose. The ability to estimate lost productivity from salaries was felt to be unreliable due to a lack of relationship between salaries and lost production. The estimation of morbidity and mortality costs resulting from suicide attempts (successful and unsuccessful) from hospital coding data was also discounted as being difficult to obtain reliably. Their estimation of £222 million direct treatment costs is much lower than that of Kind & Sorensen (1993) of approximately £416.7 million and that of the latest estimate from Thomas & Morris (2003) of approximately £369.9 million. However, these UK costs for direct treatment are much less than those for the USA as estimated by Stoudemire *et al.* (1986) and Greenberg *et al.* (1993). This may result from the population of the USA being four times greater than that of the UK, and also because of the different health delivery systems in the USA compared to the NHS in the UK.

A more recent analysis has calculated the costs of depression in adults in England during 2000 (Thomas & Morris 2003). They used a prevalence-based approach to estimate

the burden of depression. This was based on the direct treatment costs (in both primary and secondary care) and the indirect costs of lost working days (assessed as claims for incapacity benefit) and lost life years (morbidity and mortality costs). However, unlike Greenberg *et al.* who included the loss of productive capacity while remaining at work, this study did not include this estimate. Prevalence data were estimated by applying the rates of depression for 1998 from the Office of National Statistics to the population data for England in 2000. It was estimated that there were 2.6m cases of depression. 72% were female and 20% of cases fell in the 35-44 years age band. 109.7m working days were lost because of depression (Thomas & Morris 2003).

Table 8: Direct NHS Costs for Treating Depression

	Kind & Sorensen 1993	Thomas & Morris 2003
Primary care costs		
GP Consultations/visits	£126,399,000	£8,217,000
Community psychiatric nurses	£16,285,000	-
Drug Treatment	£47,280,000	£310,378,000
Social services	£40,183,000	-
Secondary care costs		
Hospital care	£165,530,000	£28,660,000
Day case	-	£476,000
Out-patient clinic	£9,146,000	£22,133,000

The total direct cost to the NHS for treating depression was estimated as nearly £370 million, of which 84% (approx £310.4 million) was attributed to antidepressant medication (Thomas & Morris 2003). These estimates are slightly lower than those calculated by Kind and Sorensen (1993). This is probably due to a change in utilisation in hospital care. The differences in direct NHS costs are described in Table 8, showing that the cost for in-patient care has reduced, although the cost of antidepressant medication in primary care has increased. This may result from the increase in the prescribing of SSRIs and other new antidepressants. Also, Thomas & Morris did not include the costs associated with community psychiatric nurses due to the lack of accurate data, nor were those stemming from the use of social services addressed. Curiously, although Thomas and Morris include inpatient care, day case, outpatient care and GP consultations, there was no mention of psychology services, which are now being promoted as first choice in primary care for mild to moderate depression (NICE 2009). The large change in GP costs appears to result, in part, from the Kind & Sorensen data including both consultations and home

visits. Surgery consultations accounted for £112,109 million, and the calculation Kind & Sorensen used a slightly higher value for the cost of a consultation (£17) than that used by Thomas & Morris (£15).

Although inspection of this data indicates medication is a high proportion of total NHS costs, when the total costs of depression (including working days lost, mortality costs) are considered, then medication is a much lower proportion: £310.4 million out of a total of over £9 billion, representing 3.4% (Thomas & Morris 2003).

This analysis has demonstrated strong evidence that depressive illness is very costly to nations in both human and societal terms. Even though there are differences in the estimates of the overall cost burden attributable to depressive illness, the trend is one of increasing cost over time. Looking at the detail within the studies for the UK shows increasing cost for drug therapy; the percentage of drug cost to total cost had trebled between 1993 and 2003. The costs of hospital inpatient care and of GP care had reduced but there was an increase in the cost of outpatient care. The reduction in hospital inpatient costs are probably due to bed reductions that have occurred in many mental health trusts during that 10-year period, while the lower spend on antidepressants in primary care may be due to increasing use of cheaper generic drugs. The increased spend in outpatients could be due to GPs referring more patients to secondary care outpatient clinics and the latter using more expensive non-generic drugs, possibly in combination to treat more resistant illness and/or at higher doses than normally prescribed by a GP. It is also interesting to note that the age of illness has reduced over the ten-year period; prevalence was 45-64 in 1993, decreasing to 35-44 in 2003.

If the new antidepressants are as effective as clinical trials suggest and the tolerability is better as the pharmaceutical companies suggest, then the question is whether the cost of the antidepressants that have been launched over the last ten years can be offset by reduction in these socioeconomic costs. The next section will review the economic studies for the antidepressants.

Economics of Antidepressants

A total of fifty-two papers were found for economic studies of the antidepressants (Table 9). Of those fifty-two studies of antidepressants, forty-two are cost-effectiveness analyses (1 citalopram, 1 duloxetine, 12 escitalopram, 2 fluoxetine, 1 fluvoxamine, 1 milnacipran, 7 mirtazapine, 3 nefazodone, 3 paroxetine, 1 sertraline, 5 SSRIs, 1 St John's Wort, 5 venlafaxine); this is the most common type of analysis. Of the remaining types of analysis, five cost-of-illness (1 citalopram, 3 fluoxetine, 1 sertraline), four cost-utility (1

bupropion, 1 imipramine, 1 nefazodone, 1 sertraline), three cost-benefit analyses (1 fluoxetine, 1 paroxetine, 1 SSRIs) and one cost-minimisation analysis (SSRIs) were identified (Table 9). The number of studies a drug has been analysed for cost implications and the types of analysis undertaken are shown in Table 10. This study will focus on the most common type of analysis, cost-effectiveness.

Most studies are funded or supported by the drug companies whose drugs are under scrutiny. At least one author of most reports was an employee, or had been a consultant for, the company. Only five studies are independent of company funding, unrestricted grants or some other form of potential influence or bias. A sixth report may be independent but no statement of funding was provided (Priest 1996). Eight other analyses either did not state details of funding sources or were not clear about such sources (Armstrong 2005; Francois 2003; Hemels 2005; Kulp 2005; LePen 1994; Lothgren 2004; Suter 2003; Thayer 2003). A few of these were conference abstracts. These short reports of work rarely seem to give any information about the source of support to carry out the study.

A large number (21% of the studies) are for escitalopram, mainly cost-effectiveness analyses. This provided further focus on escitalopram to determine why there should be so many economic studies for one drug.

Table 9: Pharmacoeconomic Analyses of Antidepressants 1980-2005

Author	Year	Country	Drug	Analysis Type	Comparator	Industry Influence
Anton	1995	Canada	Nefazodone	CEA	Imipramine, Fluoxetine	1 author
Antonuccio	1997	USA	Fluoxetine	CEA	CBT	Independent
Armstrong	2005	USA	Escitalopram	CEA	Sertraline	1 author
Bentkover	1995	USA	Paroxetine	CEA	Imipramine	Support
Borghi	2000	UK	Mirtazapine	CEA	Amitriptyline, Fluoxetine	Organon support
Boyer	1998	France	Sertraline	COI	Fluoxetine	1 author
Brown	1999a	France	Mirtazapine	CEA	Amitriptyline, Fluoxetine	1 author, support
Brown	1999b	Austria	Mirtazapine	CEA	Amitriptyline	1 author, support
Brown	2000	France	Mirtazapine	CEA	Fluoxetine	1 author, support
Dardennes	1999	France	Milnacipran	CEA	TCAs, SSRIs	1 author, support
Demyttenaere	2005	Belgium	Escitalopram	CEA	Citalopram, venlafaxine	1 author, funding
Einarson	1995	USA	Venlafaxine	CEA	TCAs, SSRIs, heterocyclics	Funding
Einarson	1997	Canada	Venlafaxine	CEA	TCAs, SSRIs	Support
Fernandez	2005	Europe	Escitalopram	CEA	Venlafaxine	1 author, funding
Forder	1996	UK	Sertraline	CEA	TCAs	Unrestricted support
Francois	2003	Norway	Escitalopram	CEA	Citalopram, venlafaxine, Fluoxetine	2 authors, funding not specified
Freeman	2000	UK	Venlafaxine	CEA	TCAs, SSRIs	1 author, unrestricted grant
Hatziandreu	1994	UK	Sertraline	CUA	Dothiepin	2 authors, support
Hemels	2004a	Austria	Escitalopram	CEA	Citalopram	1 author, funding
Hemels	2004b	Austria	Escitalopram	CEA	Citalopram	1 author, funding
Hemels	2005	Turkey	Escitalopram	CEA	Citalopram, venlafaxine	2 authors
Holm	2000	Europe	Mirtazapine	CEA	Fluoxetine	Independent
Jonsson	1994	UK	Paroxetine	CEA	Imipramine	Support
Kamlet	1995	USA	Imipramine	CUA	Interpersonal therapy	Independent
Kind	1995	UK	SSRIs	CEA	Watch-wait	Independent
Kulp	2005	Germany	Escitalopram	CEA	Venlafaxine	Not specified
Lapierre	1995	Canada	Paroxetine	CEA, CBA	Imipramine	1 author, support

Table 9: Pharmacoeconomic Analyses of Antidepressants 1980-2005 (cont'd)

Author	Year	Country	Drug	Analysis Type	Comparator	Industry Influence
Lenox-Smith	2004	UK	Venlafaxine	CEA	SSRIs, amitriptyline	2 authors, funded
LePen	1994	France	Fluoxetine	CBA	TCAs	Not clear
Lothgren	2004	Sweden	Escitalopram	CEA	Citalopram, venlafaxine	2 authors
Nuijten	1995	Germany	Citalopram	CEA	TCAs	Funding
Nuijten	1998	France	Fluvoxamine	CEA	TCAs	2 authors, grant
Peveler	2005	UK	SSRIs	RCT, CEA	TCAs, lofepramine	Independent
Priest	1996	General	Venlafaxine	CEA	Fluoxetine	Independent?
Revicki	1995	Canada	Nefazodone	CEA	Imipramine, fluoxetine	1 author, part grant
Revicki	1997b	USA	Fluoxetine	COI	TCAs	1 author, grant
Revicki	1997a	USA	Nefazodone	CEA, CUA	Imipramine, fluoxetine	Support
Romeo	2004	UK	Mirtazapine	CEA	Paroxetine	Funding
Sacristan	2000	Spain	Fluoxetine/pindolol	CEA	Fluoxetine/placebo	6 authors, grant
Sclar	1995	USA	Fluoxetine	COI	Paroxetine, sertraline	1 author, support
Sclar	1999	USA	Citalopram	COI	SSRIs, amitriptyline	2 grants
Simon	1996	USA	Fluoxetine	COI	TCAs	1 author, grant
Stewart	1994	UK	SSRIs	CMA	TCAs	Not specified
Sullivan	2004	USA	SSRIs (ESC)	CEA	SSRIs	Unrestricted grant
Suter	2003	USA	Bupropion	CUA	Sertraline	Not specified
Thayer	2003	USA	St John's Wort	CEA	Fluoxetine	Not specified
Tome	1997	UK	SSRI/augmentation	CBA, CEA	SSRI/placebo	Independent
Tome	1998	UK	SSRI/augmentation	CEA	SSRI	Support
van Baardewijk	2005	Canada	Duloxetine	CEA	Venlafaxine XR	Independent
van Loon	2002	Hungary	Mirtazapine	CEA	Fluoxetine	1 author
Wade	2005a	UK	Escitalopram	CEA	Venlafaxine, citalopram	2 authors, funded
Wade	2005b	UK	Escitalopram	CEA	Citalopram	2 authors, funded

Table 10: Number of Studies and Analysis Type

Drug	No of Studies	Type of Analysis				
		CBA	CEA	COI	CUA	CMA
Bupropion	1				1	
Citalopram	2		1	1		
Duloxetine	1		1			
Escitalopram	12		12			
Fluoxetine	6	1	2	3		
Fluvoxamine	1		1			
Imipramine	1				1	
Milnacipran	1		1			
Mirtazapine	7		7			
Nefazodone	3		3		1	
Paroxetine	2	1	3			
Sertraline	3		1	1	1	
SJW	1		1			
SSRIs	6	1	5			1
Venlafaxine	5		5			

SJW = St John's Wort
SSRIs = selective serotonin reuptake inhibitors
TCAs = Tricyclic Antidepressants

Economic Analysis of Escitalopram

Twelve economic studies of escitalopram were identified. All the studies have some degree of industry funding, with the exception of Kulp *et al.* (Kulp *et al.* 2005), which appears to be independent, although there is no indication whether any form of support was given. This study is one of three reported only as conference abstracts, and has very little data or background information to enable a quality assessment to be made. Kulp *et al.* report that the cost-effectiveness ratio for escitalopram has a 30% advantage over extended-release venlafaxine and is therefore a cost-effective alternative to the latter in the German setting. If this study did not have any industry funding, it lends veracity to there being a trend favourable to escitalopram. The studies are described in Table 11. Armstrong *et al.* (2005) is reported only as an abstract and is one of two analyses to use quality-adjusted life years to give cost/QALY as an outcome measure from the payer perspective. However there is more information than is given in the Kulp report. The data were obtained from an eight-week clinical study and published literature. The estimated six-month cost/QALY was $2362 for escitalopram and $3494 for sertraline, a saving of $1132 in favour of escitalopram. The authors suggest this advantage for escitalopram results from lower rates of adverse events and less likelihood of titrating the dose. The

third conference report is by Hemels *et al.* (2005), and compared the cost-effectiveness of escitalopram with generic citalopram and venlafaxine in a Turkish setting. There is a reasonable amount of information to inform the validity of the analysis. Again the conclusion is that escitalopram is a cost-effective treatment compared with generic citalopram and a 'cost saving' alternative to venlafaxine when treating major depressive disorder in Turkey.

Sullivan *et al.* (2004) developed a model from the managed care/payer perspective, with cost and cost-effectiveness as outcomes, using the standard six-month time horizon. The authors compared six antidepressants: citalopram, escitalopram, fluoxetine (generic), paroxetine (generic and controlled release), sertraline, and venlafaxine (standard and

Table 11 Cost-effectiveness Analysis Studies of Escitalopram

Author	Year	Setting	Comparator	Time Horizon (Months)	Outcome Measures	Success Rate (%)		Costs		Perspective	Data Sources
						ESC	Comp	ESC	Comp		
Armstrong	2005	USA	Sertraline	6	Cost/QALY			$2,362	$3,494	Payer	8wk RCT, literature
Demyttenaere	2005	Belgium	Citalopram	6	Success rate	62.3	57.2				Meta-analysis, treatment survey, literature, expert panel
					Costs:BIS			€ 390	€ 411	BIS, Society	
					:society			€ 1,162	€ 1,276		
Fernandez	2005	Europe	Venlafaxine XR	2	Cost/pt: payer			€ 110	€ 161	Healthcare payer, society	RCT, healthcare cost data
					:society			€ 765	€ 873		
François	2003	Norway (2000)	Citalopram	6	Success rate	52.2	42.4			Healthcare society	
					Costs:health			kr 6,700	kr 8,400		
					:society			kr 30,600	kr 38,000		
			Fluoxetine		Success rate		42.4				
					Costs:health				kr 8,550		
					:society				kr 38,400		
			Venlafaxine		Success rate		49.2				
					Costs:health				kr 8,200		
					:society				kr 33,800		
Hemels	2004a	Austria	Citalopram	6	Success rate	53.7	48.9				
					Costs:SHIS			€ 2,849	€ 3,803	Austrian SHIS, society	Pooled analysis, long-term trial data, switch data, literature, expert panel
					:society			€ 5,610	€ 6,979		
Hemels	2004b	Austria	Citalopram	6	Success rate	64.5	59.1			Austrian SHIS, society	Price lists, literature, RCT in primary care
					Costs:SHIS			€ 608	€ 423		
					:society			€ 3,034	€ 3,269		

92

Table 11: Cost-effectiveness Analysis Studies of Escitalopram (contd)

Author	Date	Setting (Year)	Comparator	Time Horizon (Months)	Outcome Measures	Success Rate (%) ESC	Comp	Costs ESC	Comp	Perspective	Data Sources
Hemels	2005	Turkey	Citalopram	6	Success rate	63.2	57.6				
					Costs:govern			$297	$305	Government, society	RCTs, literature, expert opinion
					:society			$678	$709		
Kulp	2005	Germany	Venlafaxine XR	70 days	Costs: GP			€ 113	€ 144	Physician	Survey of prescribers
					:specialist			€ 123	€ 163		
Lothgren	2004	Sweden	Citalopram	6	Success rate	64.9	59.3			Swedish healthcare	
					Costs:health			12,756 kr	13,871 kr		
					:society			32,533 kr	36,727 kr		
			Venlafaxine	6	Success rate	69.5	69				
					Costs:health			11,114 kr	11,489 kr		
					:society			27,735 kr	28,524 kr		
Sullivan	2004	USA	Citalopram	6	Cost			$3,891	$3,938	Managed care/payer perspective	Literature
					Effectiveness			0.341	0.340		
			Fluoxetine (generic)		Cost				$4,034		
					Effectiveness				0.335		
			Paroxetine (generic)		Cost				$4,385		
					Effectiveness				0.332		
			Paroxetine CR		Cost				$4,440		
					Effectiveness				0.332		
			Sertraline		Cost				$4,250		
					Effectiveness				0.335		
			Venlafaxine		Cost				$4,613		
					Effectiveness				0.326		
			Venlafaxine XR		Cost				$4,227		
					Effectiveness				0.336		

Table 11 Cost-effectiveness Analysis Studies of Escitalopram (contd)

Author	Date	Setting (Year)	Comparator	Time Horizon (Months)	Outcome Measures	Success Rate (%) ESC	Comp	Costs ESC	Comp	Perspective	Data Sources
Wade	2005a	UK	Citalopram	6	Success rate	63.5	58.2			NHS, society	RCTs, General Practice Research Database, expert panel
					Costs:NHS			£732	£933		
					:society			£3,635	£4,159		
			Venlafaxine	6	Success rate	68.9	68.5				
					Costs:NHS			£546	£607		
					:society			£2,640	£2,693		
Wade	2005b	UK	Citalopram	6	Success rate	53.7	48.7			NHS, society	Literature, NHS costs
					Costs:NHS			£786	£932		
					:society			£1,283	£1,521		

controlled release forms). The Sullivan study examined cost-effectiveness from the viewpoint of adverse events as the authors considered the SSRIs to be of similar efficacy. They offset the beneficial utilities arising from the efficacy of the antidepressants against adverse events utilities derived from the frequency and distribution of adverse event profiles of the individual drugs. As with Armstrong *et al.* (2005), effectiveness was estimated in QALYs, although the estimate for escitalopram was slightly lower than in the study of Armstrong *et al.* It is of note that the effectiveness estimate differences in this study and in most of the others are not large, supporting the idea that the SSRIs and venlafaxine are of similar efficacy. Escitalopram had the lowest cost ($3891: the next lowest was citalopram at $3938), although the expected cost-effectiveness is almost the same as that for citalopram (0.341 vs 0.340, respectively). The authors conclude that escitalopram is a cost-saving alternative to other SSRIs, being a dominant strategy in their model.

One study was conducted in parallel with a multinational RCT comparing escitalopram and venlafaxine and prospectively examined the costs and quality of life of 251 patients over an 8-week period from the perspectives of healthcare payer (Fernandez *et al.* 2005). EuroQOL (EQ-5D) and the Quality of Life Depression Scale were used to report patient outcomes, while the medical costs and absence from work over the previous three months measured the utilisation of medical services. Measurements were taken at baseline and at the end of the 8 weeks trial period. The effectiveness measure for the cost-effectiveness analysis was the EuroQOL score. Patients who had received escitalopram reported fewer problems on the EuroQOL score than those on venlafaxine. Payer cost was less for escitalopram than for venlafaxine XR (€110 vs €161), while the societal cost per patient was €765 for escitalopram, €873 for venlafaxine.

The remaining studies in Table 11 (Armstrong 2005, Demyttenaere 2005, Francois 2003, Hemels 2004a & 2004b, Hemels 2005, Kulp 2005, Lothgren 2004, Wade 2005a & 2005b) defined outcomes as success rates (expressed as percentage of successfully treated patients) and costs. The latter were usually from the perspective of health insurance systems, healthcare providers or society, usually from the healthcare provider aspect. Data sources were mostly from RCTs, pooled analyses, other published literature that informed the cost data, and expert opinion usually local to the country in which the study is being carried out, so that the analyses were retrospective (unlike the prospective Fernandez study). Except for the Fernandez study, all the studies were centred around decision analytic models, for which retrospective data for efficacy (ie

randomised controlled trials, resource use, and expert opinion) are key components in such modelling. The comparator drugs were all similar: citalopram and venlafaxine were most common, with one study each using fluoxetine or sertraline. Efficacy data is derived from the same trials or one meta-analysis in several of the decision analytic studies. The time-lines for the models were all six months except for Kulp who used a seventy day time-line. Two studies showed almost identical success rates for escitalopram and venlafaxine, however, although the costs were in favour of escitalopram (Lothgren *et al.* 2004; Wade *et al.* 2005a), possibly because venlafaxine usually has a higher purchase cost than escitalopram and there is the suggestion in the Lothgren study that the higher remission rate with escitalopram was a significant contributing factor. The overall result from these studies is that escitalopram is a cost-effective treatment for depression, having higher success rates than the comparator drugs and lower costs.

Table 12 examines the incremental cost effectiveness ratios (ICER) for these studies using, generally, the overall success rates from the studies for effectiveness. One study uses QALYs as its effectiveness outcome. The smaller the ICER, the greater is the cost effectiveness of the investigational drug. The impression is that escitalopram is a more cost effective drug than the comparators. All the alternative drugs are estimated to have a higher cost than escitalopram and in most instances have poorer effectiveness. However, it can be seen that venlafaxine has near-identical effectiveness in three studies (Demyttenaere 2005; Fernandez 2005; Lothgren 2004). In the Hemels (2005) study, there is insufficient detail regarding venlafaxine. This results in the ICERs for all the studies being negative: the alternative strategies are dominated.

The study by François *et al.* (2003) also supports a trend towards escitalopram being a cost-effective option in Norway, although the report is not clear. This study considered two aspects of economics. One was the impact on the Norwegian health economy of introducing escitalopram. The second was the development of a cost model that analysed the outcomes achieved at the end of six months. They included a secondary care path, which followed on from the primary care phase, where all patients started treatment. (As generally happens in clinical practice.) Data used to inform the model were derived from clinical trials of escitalopram vs. citalopram, data from published studies of fluoxetine and venlafaxine, a specific Norwegian observational study, advice from an expert panel, and 'information from the literature'. The authors calculated the average cost-effectiveness ratio to give the expected cost per successfully treated patient. The abstract (from which the data in Table 12 are taken) reports the average expected

total costs per patient but the estimation of these is not stated in the article itself. As no head-to-head trials to test the four drugs examined existed, data for fluoxetine and venlafaxine were imputed from randomised controlled trials data that used a common reference.

Table 12: Escitalopram Cost-effectiveness Ratios

Author	Drug	Cost (C)	Effectiveness (E)	Incremental Cost (IC)	Incremental Effectiveness (IE)	ICER (IC/IE)
Armstrong	ESC	$952	0.403			
2005	SRT	$1,372	0.393	€ 420	-0.01	-42000
Dettymeare	ESC	€ 626	62.3			
2005	CIT	€ 719	57.2	€ 93	-5.1	-18
	ESC	€ 497	67			
	VNF	€ 525	66.6	€ 28	-0.4	-70
Fernandez	ESC	€ 110	0.78			
2005	VNF	€ 161	0.77	€ 51	-0.01	-5100
Francois	ESC	kr 19,661	64.2			
2003	CIT	kr 22,379	58.7	kr 2,718	-5.500	-494
	FLX	kr 22,558	58.7	kr 2,897	-5.500	-527
	VNF	kr 20,989	62.1	kr 1,328	-2.100	-632
Hemels	ESC	€ 1,547	53.7			
2004a	CIT	€ 1,851	48.7	€ 304	-5	-60.8
Hemels	ESC	€ 392	52.1			
2004b	CIT	€ 427	42.8	€ 35	-9.3	-4
Hemels	ESC	$297	63.2			
2005	CIT	$305	57.6	€ 8	-5.6	-1
	VNF	not specified				
Lothgren	ESC	SEK 12,756	64.9			
2004	CIT	SEK 13,871	59.3	€ 1,115	-5.6	-199
	ESC	SEK 11,114	69.5			
	VNF	SEK 11,489	69	€ 375	-0.5	-750
Sullivan	ESC	$3,891	0.341			
2004	CIT	$3,938	0.340	$47	-0.001	-47,000
	FLX	$4,034	0.335	$143	-0.006	-23,833
	PRX	$4,385	0.332	$494	-0.009	-54,889
	PRX CR	$4,440	0.332	$549	-0.009	-61,000
	SRT	$4,250	0.335	$359	-0.006	-59,833
	VNF	$4,613	0.326	$722	-0.015	-48,133
	VNF XR	$4,227	0.336	$336	-0.005	-67,200
Wade	ESC	£465	52.8			
2005a	CIT	£544	43.5	€ 79	-9.3	-8
	ESC	£374	69.9			
	CIT	£413	69.7	€ 39	-0.2	-195
Wade	ESC	£422	53.7			
2005b	CIT	£454	48.7	€ 32	-5	-6.4

CIT = citalopram; ESC = escitalopram; FLX = fluoxetine; PRX = paroxetine; SRT = sertraline; VNF = venlafaxine; CR = controlled release; kr = Norwegian kroner; SEK = Swedish kroner.

Citalopram was used for the fluoxetine comparison while for venlafaxine, fluoxetine was the common reference (derived from RCTs of venlafaxine vs fluoxetine).

There are comparative data between escitalopram and venlafaxine since the publication of the François *et al.* paper (Montgomery *et al.* 2004; Bielski *et al.* 2004). However, François has been an author in a recent paper (Fernandez *et al.* 2005) that estimated the cost-effectiveness of escitalopram versus venlafaxine in a study conducted alongside a European double-blind, randomised controlled trial. Unlike the majority of cost-effectiveness analyses, this was to be a prospective analysis. The two drugs were of similar effectiveness, measured in terms of quality of life (EQ-5D and QLDS). However, there is a price difference between the two drugs: at the 2003 euro values used in the study, average cost for escitalopram was €0.86 and €1.08 for venlafaxine.

The evidence therefore suggests that escitalopram is a cost-effective option in several health economies. However, from a cost viewpoint, as the comparator drugs are considered of equal efficacy and probably effectiveness, the cheapest drug should be considered. Lundbeck have promoted escitalopram as a cost-effective option to venlafaxine. If the cost of the latter is more than that of escitalopram, then the Lundbeck product should be considered.

Cost Models

An early example of such a model was that developed to evaluate episodic antidepressant drug therapy in a primary care setting, using either imipramine or amitriptyline as the TCA, or sertraline or paroxetine as the SSRI (Stewart 1994). The alternative drug regimens were compared in terms of expected total costs per patient (i.e. drug costs and other health care). Although described as a cost-minimisation analysis by the authors, the analysis performed was a retrospective cost-effectiveness analysis as the cost per successfully treated patient was calculated. The cost per treatment was summed and the average used to calculate an average cost-effectiveness ratio (ACER). This is no longer a recommended method. It was assumed the patient allocation to a particular treatment would be random, although in real-life there may well be selection bias by the GP. Their decision tree does not appear to allow for dose escalation, a not uncommon event. The study was also designed around the idea of episodic treatment, whereas this may not be a true representation of actual clinical practice. The results showed that, although the cost per successfully treated patient were not as low for TCAs as had been

expected (imipramine £491; amitriptyline £539; sertraline £581; paroxetine £547), there was not a clear cost argument in favour of switching from TCAS to SSRIs.

A similar approach was used by Jonsson and Bebbington (1994) who constructed a decision tree to provide a model of clinical practice. As with Stewart (1994), allowance was made for relapse and switching between drug therapies. Much of the data used to construct the tree was derived from RCTs, but an expert panel was also used 'to bridge the inevitable gaps' in the model and for advice about patient management and treatment patterns. Expert estimates of resource use and probabilities can over-estimate real world costs and situations (Simon *et al.* 1995). Not all costs were included in the model: e.g. those associated with suicide or the cost of doing a home visit. A pooled analysis of 726 patients by Dunbar *et al.* (1991) was used to estimate the drop-out rate between the two drugs and it was assumed that the drop-out rate for paroxetine was lower than that for imipramine. As there was no difference in efficacy between paroxetine and imipramine in compliant patients, failure to respond to treatment was not included in the model. Patients were assumed to receive the drugs for twelve weeks, although the Dunbar analysis looked at 6-week trial data. Compliance therefore determined efficacy. Again, their results showed a marginal difference in expected costs per patient irrespective of outcome (£430 for paroxetine, £424 for imipramine) but, for the costs per successfully treated patient per year, the paroxetine cost was lower than that for imipramine (£824 vs £1024). Sensitivity analysis showed that this difference was reduced if the paroxetine was not as effective, but it still had a lower cost per successfully treated patient. Like that of Stewart, the model is simplistic and does not allow for all real-life clinical situations. It is also of note that this work was supported by SmithKline Beecham (now GlaxoSmithKline), the company that manufactures paroxetine.

With regard to escitalopram, Figure 9 (reproduced from Hemels *et al.* 2004a, with permission of the TheAnnals.com) shows the core part of a decision tree that modelled the cost-effectiveness of escitalopram versus citalopram from the perspectives of Austrian society and the Austrian Social Healthcare Insurance System. Effectiveness was defined in terms of remission of symptoms (MADRS score <=12) at six months. Costs were calculated from the resources used along each treatment path. This generated an incremental cost-effectiveness ratio (ICER) expressed as the expected additional cost per patient in symptom remission. The model describes the paths from the initial choice of antidepressant, the possible outcomes of each choice at each node, and the probabilities associated with each node branch. The probabilities were based on remission rates

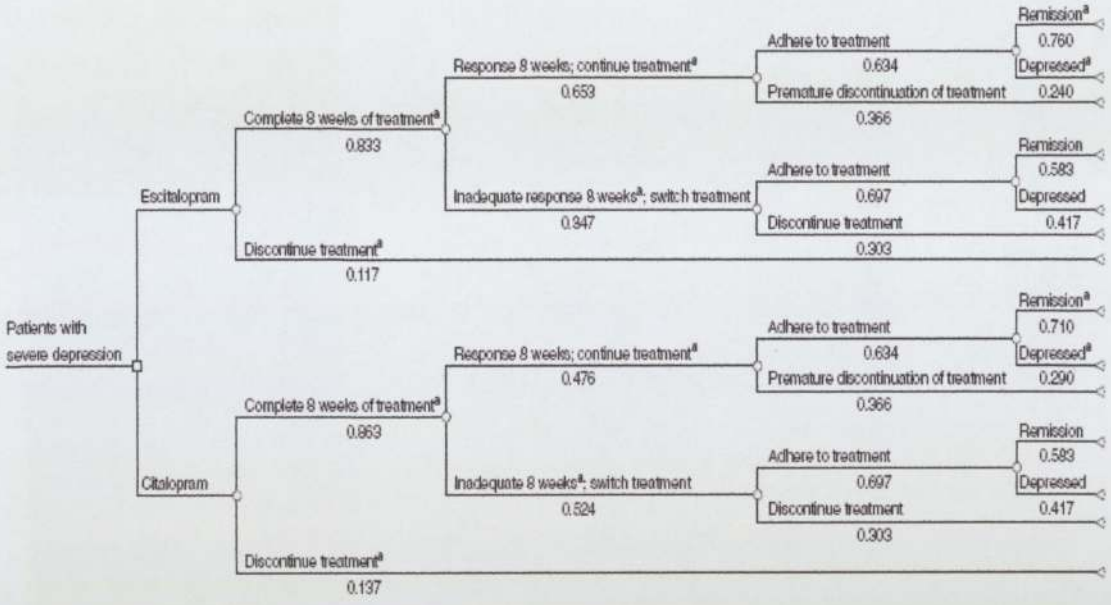


Fig 9: Decision Analytic Model for Escitalopram vs Citalopram

derived from earlier studies, and the model was scrutinised by a Delphi panel to ensure its applicability to the management of severe depression in Austria. The analysis also included two other models: discontinuation and suicide.

Real World Trials: an alternative

In an attempt to avoid some of the problems associated with randomised controlled trials and the more theoretical economic approach of modelling, Simon *et al.* (1995; 1996) suggested a hybrid trial model: the real-world randomised trial. This would apply the principles of RCTs to real-life studies. The concept involves a prospective longitudinal design and would determine outcomes in terms of effectiveness and cost-effectiveness. RCTs only look at efficacy under strictly controlled conditions.

Simon *et al.* proposed using initial randomisation with subsequent unblinded reassignment to alternative medication: this would follow more closely what happens in real life clinical activity when switching. The investigator would be blinded to patient assessments by second researchers using telephone interviews with patients: i.e. data collection is separated from clinical management. It would also reduce the loss of data that would result from patients not attending clinics. The integration of some of the principles of RCTs into observational 'real world' yields three prime benefits: 1) important outcomes can be accurately and relatively unobtrusively measured; 2) the impact of depression and its treatment on healthcare utilisation can be measured quite accurately; 3) cost-effectiveness analyses are based on the costs of care delivery.

Furthermore, although treatment is initiated with one drug, which may later be switched to a different one, the analysis is based on the initial treatment.

The authors have reported such a trial (Simon *et al.* 1999). The only controlled aspect was the initial treatment prescribed, which was randomly assigned to desipramine, fluoxetine or imipramine. Neither physicians nor patients were blinded to this initial treatment, in order to mirror as closely as possible clinical practice. Most assessments were carried out over the 24-month trial period by telephone interview and included depression rating scales and quality of life instruments. To determine the most appropriate choice of initial antidepressant, an intention to treat method was used based on the initial study medication.

However, initial treatment assignment did not preclude switching to a second antidepressant, which reflects what may happen in normal clinical practice. The proportion of patients continuing their original antidepressant treatment decreased over time, regardless of the drug used initially, but the likelihood of patients continuing with their initial antidepressant was greater in the fluoxetine group. When switching occurred, more than 60% of medication switching or discontinuation during follow-up was observed during the first six months, and was less likely with fluoxetine.

Approximately 35% of patients who started treatment with TCAs switched to alternative antidepressants at some point during follow-up. These switches may have implications for clinical outcomes and treatment costs.

The greatest improvements in both symptoms and quality of life were seen in the first six months, with a slowing of improvement after that although statistically significant improvement continued, but achievement of remission was seen in less than half the patients at each assessment, regardless of the treatment arm. Efficacy measures (HAMD and SF-36) were not clinically significant. Also, those patients who switched experienced a delay in recovery, while those who continued with the original treatment did not. Simon *et al.* also considered total medical costs for the interventions. Costs were found to be the same for all three groups. In comparison to total medical costs, the drug costs were found to be relatively small. Fluoxetine did not reduce total medical costs. This suggests that the drug costs do not seem to influence the total cost of treatment, at least in this setting.

This confirms the authors' earlier, preliminary findings (Simon *et al.* 1996) and is supported by similar findings by Woods and Rizzo (1997) who found that there is no cost

advantage using SSRIs first, but contradicts other findings from decision analytic models (for example, Lapierre *et al.* 1995; Jönsson & Bebbington 1994).

A similar trial has been conducted more recently by Kroenke *et al.* (2001). This was an open-label, randomised, intention to treat trial comparing the effectiveness of paroxetine, fluoxetine and sertraline. 573 patients were entered, randomly assigned to one of the antidepressants. Following initial randomisation, primary care physicians were then free to switch patients to a different SSRI and even non-SSRI if they did not adequately respond to or tolerate the initial SSRI. As with the Simon *et al.* trial (1999), randomisation was not blinded, although allocation occurred after the patient had seen the physician. (This is not clear in the Simon paper.) Structured telephone interviews were used to collect data, which were not disclosed to the physician. Unlike the Simon *et al.* study, Kroenke *et al.* did not use any depression rating scale (eg HAMD or MADRS) but used other inventories, the primary one being the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) Mental Component Summary. This instrument incorporates all eight SF-36 subscales as a measure of mental health in a regression algorithm. The authors claim it 'has been established as a sensitive outcome measure in studies of clinical depression' (Kroenke *et al.* 2001). In addition, the Symptoms Checklist (SCL-20) was used as it has been shown to detect differences between treatment groups in primary care settings with sufficient sensitivity. Social function and work function were evaluated using a battery of six measures. The lack of inclusion of a depression rating scale could be a drawback to this study, as there would be no indication of the severity or depth of depression.

The study demonstrated the lack of differences between the three SSRIs across a broad range of outcomes over the 9-month period of the trial. Loss of patients to follow-up was low, such that the number of patients at nine months meant the study still had 94% power to detect a difference of five points on the main outcome measure. (A 5-point difference on this scale represents a half a SD, which is a medium effect size.) There were no significant differences in reasons for discontinuation between the three SSRIs. One confounder, however, is that patients did not have to pay for medication, which is not usual practice. This may have encouraged better compliance/adherence, although this situation mirrors that in the NHS for those patients exempt from paying prescription charges and adherence is still a problem in the UK. This may be due to other factors, such as education about the treatment (Masand 2003), social class, or stigma (Dinos *et al.* 2004; Roeloffs *et al.* 2003). A commonly prescribed antidepressant in the UK and

Europe (François, 2003; Lothgren, 2004), citalopram, was not included in the trial. The overall conclusion of the study was that none of the three antidepressants has superiority over the other two.

Discussion

It is only in recent years that valid research has been conducted on the cost and impact of mood disorders on the economy of a nation.

The current study has reviewed two aspects of drug economics, namely the cost of depressive illness and the costs associated with its pharmacological treatment. Treatment costs have been estimated by looking at the acquisition costs, which are retrospective using expenditure data. However, using cost models in the form of decision analytic models enables an estimate of future costs to be made. Using these techniques allows for changing the probability variables for various branches of the decision tree: these might include acquisition cost, the level of morbidity or mortality, the likelihood of discontinuation, or switching from one treatment intervention to another. Such sensitivity analyses can show whether changes in parameters affect outcomes and costs.

This chapter shows that the acquisition cost of drugs is not the only important cost involved with treating or failing to treat depression. This is only one component of direct costs: others being the time taken to diagnose the illness, out-patient or other day care (including rehabilitation). Indirect costs arise from the loss of productivity through absenteeism or presenteeism of the patient, or because the family takes time out to care for them. Such costs are more difficult to estimate and are usually valued in terms of salary but are generally considered to be greater than the direct costs, particularly drugs.

Modelling can help determine the resource allocation by estimating the costs associated with a particular course of action, and identifying the major factors and parameters of interest. The model must, however, reflect and describe the process under study accurately but most modelling so far has ignored some aspects of the process in order to simplify the model and its outcomes. Allowance for this, along with robust data input, will determine the accuracy of the model. Much of the data used will be from RCTs, which tend to be short-term and so rigidly controlled as to rarely reflect actual clinical practice. The expert opinion that is often used in constructing decision tree parameters to inform probabilities of an event occurring must ensure that the opinion is that of the psychiatric community at large.

Retrospective claims data have been used to develop cost analyses but these have drawbacks. Although the cost data is usually accurate and robust, there is often no information on disease severity or clinical outcome. In addition, the less tangible costs associated with functional impairment and work productivity are not included (Berndt *et al.* 2000).

Unlike Australia, where the industry has to provide pharmacoeconomic data to the regulatory body before approval may be granted (Henry 1992b; Judith Longworth, personal communication), Britain does not yet require such data for approval of a licence application. Meta-analysis of several trials will help to address this post-marketing, but ideally prospective pharmacoeconomic trials should be carried out. Post-marketing studies are most likely to be carried out by the pharmaceutical company. Better still, such studies, as well as efficacy studies, should be performed in the real world, an idea that will have its detractors due to the lack of rigour potentially inherent in such trials. The uncontrolled populations in such studies will have medical and psychiatric comorbidity, bringing with it additional variation that reduces the statistical precision for a given sample size. Unobserved variables account for this variation. For example, the previous psychiatric history and response to antidepressants will influence the choice of initial antidepressant for the current episode, and subsequent choice if a switch is required. The characteristics of the physician, such as attitude towards patients with mental illness, prescribing preferences and previous experience with antidepressants (most doctors have a particular first-line choice), may also influence the choice of initial antidepressant. The lower level of intervention in real world trials and the 'treatment as usual' intervention for the control group, compared with the intense scrutiny of subjects in controlled trials and while being closer to what happens in clinical practice, does mean that intervention effects are less likely to be large (Sturm *et al.* 1999). This would mean large sample sizes in each arm of the trial would be needed: even the Simon *et al.* (1995) total sample size of 536 patients cannot be considered large (Sturm *et al.* 1999). The reasons are threefold. First, that the distributions of several measures, such as costs, are very skewed and do not have upper limits. Second, sample heterogeneity will have an effect on these measures, increasing outcome measure variance and possibly reducing effect sizes. Third, meaningful change will differ according to the outcome measured: the percentage change appropriate for a quality of care measure may represent a small effect, but the same percentage change for health care costs may be dramatic.

Such factors, if not accounted for in analysis, can lead to biases in the estimates of treatment outcomes. In an RCT, these can be evened out across the arms of the study and by having rigid entry criteria. However, in retrospective or real world studies this may be less easy. There are three methods are available to correct for these inherent biases in such trials.

In the first two approaches, instrumental variables and parametric sample selection, variables are specified that correlate with the treatment selection but are uncorrelated with outcomes. In the former method, 'an instrumental variable is one that has the characteristic of being highly correlated with the variable for which it is intended to serve as an instrument without it being correlated with the error terms' (Crown 2001). If selection into treatment groups is not random, then the statistical model will contain an error of missing variables measurement. Therefore if there is correlation between the error term of the drug selection equation and treatment outcomes, treatment effect estimates will be biased. A search of PubMed reveals that there appears to have been no studies in depression using this technique.

The second technique available to overcome biases in observational studies is parametric sample selection, which is conducted as a two-stage process. Firstly, an adjustment factor, λ, is calculated for each patient, based on an estimated model of treatment selection. This factor is constructed from the errors in correctly predicting treatment selection, and then used in the second stage as an explanatory variable in the outcome model. The adjustment factor can indicate if selection bias is present if its coefficient in the outcome equation is large. In that case, the treatment effect would have been biased without such adjustment. This technique has been used in some analyses (Crown *et al.* 1998a; 1998b; Hylan *et al.* 1998).

Unlike the instrumental variables and parametric sample selection methods, propensity score analysis does not require the identification of variables that correlate with treatment selection but not with outcomes. Like sample selection, the conditional probability of a treatment outcome is first estimated for each patient. Patients are then grouped into similar probability score bands: these are the propensity scores. Finally, each group of similar propensity scores is evaluated for treatment outcomes. When the group estimates have been calculated, they are combined to give an overall treatment effect. Unlike parametric sample selection, propensity score analysis does not detect selection bias.

However, notwithstanding these difficulties, such studies have been performed (Kroenke *et al.* 2001; Simon *et al.* 1999). Both used MOS SF-36 and the Hopkins Symptoms Checklist as the main outcome measures, although only Simon *et al.* used the Hamilton Rating Scale for Depression. They do not appear to have corrected for biases using the techniques described above. This type of trial should have a measurement of the severity of depression but it is equally important to assess the social and productivity functions of patients. More trials of this type need to be conducted. Real world patients could be entered into a longer-term post-marketing surveillance study that captures data for effectiveness and quality of life, social functioning and productivity.

Pharmacoeconomic studies in depression are usually observational and retrospective in nature and therefore investigators have little or no control over treatment assignment, creating potential for bias in outcomes. To overcome these biases, two main strategies are current. One is to use a semi-randomised and controlled trial method, such as that employed by Simon *et al.* The other is to use statistical techniques to compensate for these biases. To date, it appears that neither has been extensively studied. A combination of the two may provide an even better understanding of the role of newer antidepressants in the real world, by employing a more rigorous approach to allocation of drug therapy and a statistical approach to reduce the bias inherent in observational studies.

The traditional outcomes used for these analyses are based on long-term end-points; clinical outcomes which are intermediate (eg glucose levels, blood pressure, depression scale ratings) are not appropriate unless unavoidable. However, work by Caro *et al.* has shown it is possible to use specific intermediate outcomes to determine costs associated with them in life-long diseases such as diabetes (Caro *et al.* 2004) in which specific markers (in this case, HbA_{1c} and post-prandial glucose) can be used as continuing outcome measures. Whether such an approach can be used for mental health economic analyses is not so clear as the rating scale measurements are more subjective than glucose levels. A stable glucose or HbA1c level infers the patient is in remission. The HbA1c is a more stable marker of glucose control than measuring glucose levels and therefore can perhaps demonstrate remission. Such an inference cannot be made from trials usually lasting 6-8 weeks only, the end-points of which usually demonstrating short-term efficacy rather than long-term remission. Economic analyses could be included as part of RCTs, but the carefully controlled conditions of RCTs do not generalise to the normal practice in the community. The populations in RCTs are usually

regarded as being sub-populations of the general population, so the use of cost-effectiveness data derived from these trials also may not be generaliseable. In addition, larger sample sizes are required for a period longer than that usually seen in RCTs to reduce uncertainty in cost effectiveness estimates.

If life-years gained are not the final outcome in depression trials (usually outcomes are response and remission) then the number of patients in remission could be used, but it would need to be shown that this equates to a positive health-state over time. This study has not revealed such a link. Ideally, cost-effectiveness studies should be carried out prospectively, and should study the effectiveness of a treatment (i.e. outcomes achieved in the real world) as opposed to its efficacy (which results from outcomes in a randomised controlled clinical trial). If efficacy must be used for cost-effectiveness analysis, then adjustments should be made to convert the results into real world outcomes, if that is possible. Analyses should be clear that they refer to either cost-effectiveness (real world) or cost-efficacy (clinical trial) data. This study shows that as a marker for effectiveness, QALYs have been used; there will be an improvement in the quality of life scores if the drug is effective in treating depression with good tolerability.

The decision analytic models reviewed here usually take into account the possibility of suicide. Suicide is very costly as it will often involve at least one of the emergency services and emergency health services. Such attempts can be the result of the illness itself, although antidepressants have also been implicated in the exacerbation or development of suicidality, and may have been used as part of the suicide attempt which, before the advent of the SSRIs, may have led to successful suicides.

If the probability of a successful suicide is lower with the newer drugs, this would be a more favourable outcome in the decision tree. However, it is likely that the risk of suicidality would be the same across comparator drugs, and the seven of the twelve cost-effectiveness studies evaluated here account for suicide by assuming it would be the same across all drugs. The remaining five studies do not appear to have taken suicide into account. However, although the use of emergency services might the same, due to the lower risk of cardiotoxicity with SSRIs, there is the possibility that hospitalisation might be of shorter duration and therefore less costly.

It thus seems likely that antidepressants can offset the economic, as well as clinical and social, burden of depression. However more investigation needs to be performed.

Out of the fifty-two pharmacoeconomics studies found, a seemingly disproportionate number are for escitalopram: a total of twelve, all of them cost-effectiveness analyses. Given that Lundbeck was accused of bringing out a product similar to one it already had on the market in a less than appropriate fashion (Dyer 2003), the number of pharmacoeconomic studies for escitalopram may give the impression that the company has a point to prove. This was further compounded by the company, when distributing the available literature, only having reproductions of posters from conferences. Although possibly useful indicators of a drug's potential, posters do not always give sufficient data for further analysis to be undertaken. Reliance was placed on pooled analyses, posters and brief conference proceedings reports to convince practitioners of the merits of escitalopram, as well as papers by Sanchez and her co-workers describing the pharmacology supporting the drug. This will be further explored in Chapter 5, which examines the introduction of escitalopram.

CHAPTER 5

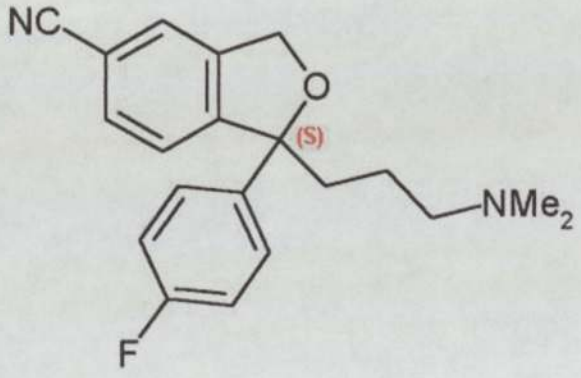
META-ANALYSIS OF ESCITALOPRAM

Introduction

Escitalopram was brought to market as the patent for citalopram was about to expire, and Lundbeck claimed that it had at least the same efficacy as citalopram. It was also claimed that the side-effect profile of escitalopram would be better. Further it was suggested that the onset of action of escitalopram was faster, an assertion partly based on the pre-clinical animal pharmacology (Montgomery *et al.* 2001). The question arose in pharmacy circles as to whether the product should be included in health economy formularies as there appeared to be no added benefit to justify the extra cost above citalopram.

Escitalopram (ESC) is the S-enantiomer of citalopram (Figure 10). Since ESC is one-half of the racemic mixture that comprises citalopram and shown in animal experiments to be the active enantiomer, it should follow that ESC should be at least as efficacious as CIT and with at least the same good side-effect profile.

Figure 10. Structure of escitalopram



Most marketed antidepressant drugs are chiral in nature: i.e. they exist in two asymmetrical forms that are mirror images of each other (Lane & Baker 1999), designated R and S enantiomers according to their ability to rotate polarised light. Each enantiomer potentially has different pharmacological properties, one usually having either greater efficacy or potential to cause side effects (Tucker 2000). The racemate may have more complex pharmacology or kinetics. However, an increasing number of drugs are being marketed as single enantiomers, often as a product extension when the patent is about to expire.

Using a single enantiomer instead of a racemate may therefore simplify dose-response relationships and, as in the case of escitalopram, reduce the total amount of drug

required to bring about a therapeutic effect and reduce potential for antagonistic effects to the desired pharmacological effect.

Although one of the isomers may have improved pharmacological properties while the other may cause some of the adverse effects, and development of the former may be claimed to be advantageous as with, for example, esomeprazole (Andersson 2004), there is evidence to suggest that stereoselective isomers may be not be of huge benefit to patients, at least with proton-pump inhibitors (PPIs) (Kromer, 2001). However, the pharmacodynamics and pharmacology of escitalopram are different to PPIs, and there is strong evidence from animal studies to suggest the S enantiomer of citalopram has advantages over the racemate or the R enantiomer. This does not necessarily mean that it should be included in health economy formularies.

This Chapter will review the pre-clinical data to assess the validity of the claim for faster onset of action and lower dose than citalopram, the efficacy data from RCTs, and will present a meta-analysis of the published primary data with a comparison of the results with those of another independent meta-analysis addressing the efficacy of escitalopram.

Method

Most of the pre-clinical data and early trial literature was made available by the local hospital pharmaceutical representative from Lundbeck at the time of the product's launch. Following this, systematic searches were made using PubMed and World of Science databases between the years 2000 to 2005 using the keywords escitalopram, citalopram, serotonin reuptake transporter, enantiomer, and randomised clinical trial to locate other trials and pre-clinical papers not provided by the company. The Medical Information department at Lundbeck was telephoned or emailed to request papers from these searches. The reference lists in the papers obtained were inspected to find any trials previously not found.

All the available trials were reviewed qualitatively and quantitatively. An early attempt was made to obtain the data from the failed trial, MD-02, but Lundbeck only sent a brief summary of four trials from a Swedish review, one of which was this study (Personal email communication from David Simpson, Lundbeck, 4th March 2004). However, a later search of the FDA website revealed a summary of the 'failed' MD-02 trial.

From each trial, data were abstracted on response (defined as a reduction of MADRS score from baseline of $\geq$ 50%), time to response, decrease in MADRS score from baseline to endpoint and the associated SD or SE, difference in end point scores between

ESC and CIT and/or placebo, and number of drop-outs. The primary end point was noted and the discussion checked to see if this was reported or whether the secondary end-point was used instead. All statistical analyses for the meta-analysis were carried out using StatsDirect Statistical Software, version 2.5.5 (StatsDirect Ltd, Sale, UK).

The trials were qualitatively reviewed using the CONSORT statement (Moher *et al.* 2001) to check that the study reports conformed to internationally accepted methods.

Statistical methods

In order to conduct a meta-analysis for this study, the identified randomised controlled trials were scrutinised for quality of reporting according the CONSORT criteria. Each paper was then analysed for homogeneity in certain parameters: study populations, duration of illness, the primary outcome measure used, the degree of severity of illness (baseline MADRS score), the study inclusion and exclusion criteria, and similar definitions of response and remission.

The meta-analysis is a quantitative review of the RCTs, bringing together the results of the individual trials, which helps inform the size of effect of using escitalopram. The individual effect sizes are statistically combined, weighting the studies according to the study size. The outcome measures must all be the same and the studies must all report the same outcomes. Pre-defined criteria, as listed above, are essential to find appropriate studies that have data that are comparable and can be pooled. The quality of the original trials determines the robustness of the meta-analysis and a quality analysis of trials should be performed. In this analysis, the CONSORT criteria will be used (Moher *et al.* 2001). In a fixed effects model, the results of an analysis are conditional on the populations of the included studies. It assumes that all estimates are of the same parameter value (e.g. mean). A random effects model does not assume this but instead considers that each sub-population has its own parameter value. Therefore the variance estimates include two levels of variability; the intra-sub-population variance and the inter-sub-population variance.

The results are described as Forest plots, which show the data for both individual studies and the combined result. The point of no difference between test and standard is drawn at 1. Weighting of studies is shown as boxes of varying size (according to the size of the study) with confidence intervals shown as horizontal lines. The combined effect is shown as a diamond shape with a vertical line describing its position in relation to the individual studies and the line of no difference, with a horizontal one showing the

confidence interval. Any confidence intervals touching or crossing the line of no difference means that there is no difference between active and comparator.

The Q statistic will be used to test for homogeneity. It is also essential to ensure that duplicate reports are not included in pooling: studies may be reported in more than one publication, either as identical or near-identical reports or with minor modifications which do not actually add more independent data.

Outcome Variables

The studies of escitalopram analysed here include two main types of outcome variables. One is dichotomous: response is given by a cut-off point on the MADRS scale of a decrease in score >50% from baseline. A subject either achieves this or does not. Conversely, the MADRS scores themselves are continuous variables until they reach the response point. Time to response is also a continuous outcome.

To summarise the dichotomous outcomes, the odds and risk ratios are pooled. An odds ratio (OR) describes the ratio of odds for the event of interest in the test group relative to the control group. This analysis attempts to describe the likelihood of response with escitalopram versus the placebo or comparator. The risk ratio describes the ratio of the probability of responding to the test drug (escitalopram) versus the control group:

$$\text{Risk ratio (RR)} = \frac{\text{probability of response in treatment group}}{\text{Probability of response in control group.}}$$

To calculate the OR and RR for a given study, a 2x2 contingency table is constructed, as in Table 14.

Table 14: Contingency table for representing the results of a trial

	Treatment	
	Escitalopram	Comparator
Responders	a	b
Non-responders	c	d
Total	a+c	b+d

The odds ratio and risk ratio would be given by

$$\text{OR} = \text{odds ratio} = \frac{a/c}{b/d} \qquad \text{RR} = \text{risk ratio} = \frac{a/(a+c)}{b/(b+d)}.$$

The risk difference is the difference in proportion of responders between escitalopram and the comparator

RD = risk or rate difference = b/(b+d) - a/(a+c).

To estimate the precision of the odds ratio, the confidence interval (CI) around each point estimate is calculated. The usual estimation is the 95% CI which, on repeated calculation, will usually include the true population mean 95% of the time. Confidence intervals define the values within which the differences in response between groups may fall.

To pool the odds ratios in a fixed effects analysis, each estimate of effect has a weighting assigned to it, ($\ln(OR_i)$ in this case). This gives studies with greater numbers of subjects more weight than smaller studies.

Heterogeneity

Fixed-effect analyses assume that the effect estimated is the same throughout between the studies. This assumption needs to be confirmed by using a chi^2 test of heterogeneity, often denoted by Q, based on a weighted sum of the squares of the differences between the log odds ratios estimated from the individual studies and the summary log odds ratio. The larger the value of Q, the more likely is it that the effects between the studies differ. The interstudy variance using the method developed by DerSimonian and Laird (1986) can be estimated by calculating V_I from Q and the mean of the weights.

A more recent measurement of heterogeneity in meta-analyses is the quantity, I^2, which measures the degree of inconsistency in results of studies (Higgins & Thompson 2002; Higgins *et al.* 2003). I^2 describes the total variation across studies due to heterogeneity rather than chance in terms of percentage. If Q is the heterogeneity statistic and df the degrees of freedom, then

$$I^2 = 100\% \times (Q-df)/Q.$$

Negative values are put equal to zero, as I^2 lies between 0% and 100%. If there is no observed heterogeneity, I^2 will be 0%, while larger values indicate increasing heterogeneity. I^2 values of 25%, 50%, and 75% have been assigned descriptions of low, moderate and high. Advantages of I^2 are that it does not inherently depend on the number of studies in the meta-analysis, nor does it depend on the outcome measure (eg odds ratio) or the data type (eg dichotomous).

Results

Pharmacology

Citalopram is one of the most selective SSRIs (Hyttel 1994), inhibiting the serotonin reuptake transporter (SERT) in the pre-synaptic terminal of serotonergic neurons. It has been shown to be a potent antidepressant in trials, with a good safety and side-effect profile (Bouchard *et al.* 1987; Ekselius *et al.* 1997; Patris *et al.* 1996; Rosenberg *et al.* 1994; Shaw *et al.* 1986; Stahl *et al.* 1998), including the elderly (Elsborg 1991). Escitalopram is the S-enantiomer of citalopram, which is an equimolar racemic mixture of

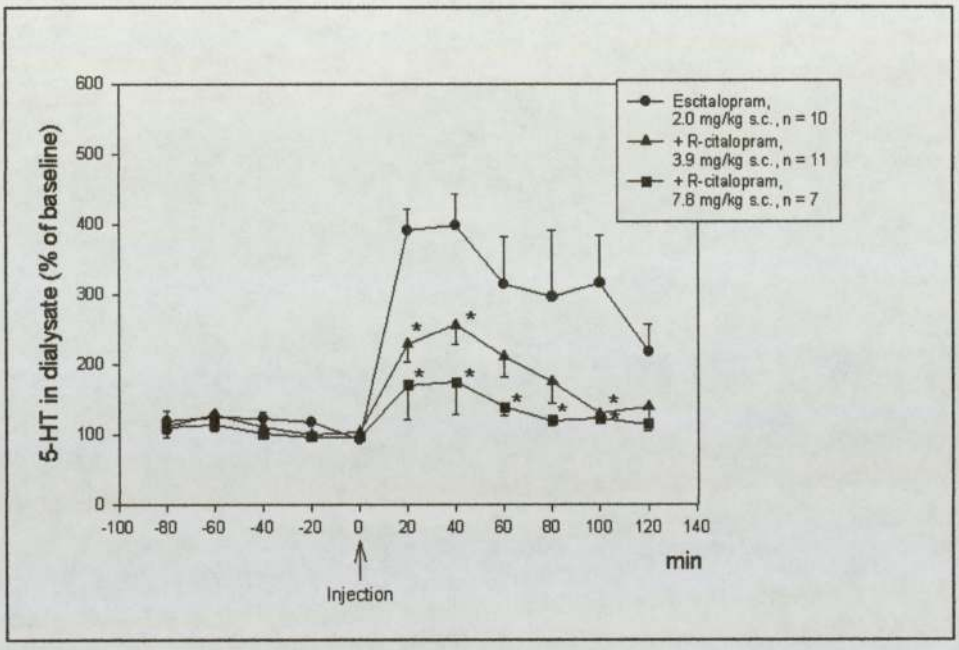


Fig. 11. Effect of R-citalopram on the escitalopram-induced increase in extracellular levels of 5-HT in the frontal cortex. Asterisks indicate values that differ significantly from the corresponding values with escitalopram alone (Tukey test: $P<0.05$). Reproduced with permission from Dr A Mørk. (Mørk *et al.* 2003)

R- and S-citalopram. Escitalopram was selected for further development as it was found to possess the inhibitory activity at the SERT whereas the R-enantiomer did not (Sorbera *et al.* 2001). Indeed, both in-vivo and in-vitro data suggest that the R-enantiomer may inhibit the effect of the S-enantiomer (Mørk & Sanchez 2003 Mørk *et al.* 2003; Sánchez 2003; Sánchez *et al.* 2003a), and that ESC is twice as potent as the racemate, confirmed by experiments using cultured cells expressing the human transporter protein (Owens *et al.* 2001; Sánchez *et al.* 2003b). This is used to support the claim that escitalopram has a faster onset of action than citalopram (Montgomery *et al.* 2001). In addition, escitalopram is at least 27 times more potent than R-CIT as an inhibitor of transporter activity in cultured cells expressing the human transporter protein (Sánchez *et al.* 2004). Microdialysis

measurements of levels of extracellular serotonin in the frontal cortex of rats after acute administration of CIT, ESC and R-CIT showed that 2mg/kg ESC elicited a 2-fold increase in brain serotonin levels when compared to 4mg/kg CIT (Mørk & Sanchez 2003). Administering R-CIT and ESC together in a 2:1 or 4:1 ratio caused a significant dose-dependent reduction in serotonin levels (Fig 11). This becomes relevant when the metabolism in humans is considered, as the two enantiomers are eliminated differentially giving a steady state R:S ratio of 2:1 (ie in favour of R) (Foglia 1997; Sidhu 1997; Brøsen & Naranjo 2001).

Based on earlier experiments that demonstrated that measurements of serum corticosterone levels could be used as a functional marker for serotonergic neurotransmission activity (Fuller *et al.* 1996; Attenburrow *et al.* 2001), other experiments where the proportion of R-CIT was increased in a mixture of R-CIT and ESC have shown decreased serotonergic neurotransmission activity (Sánchez & Kreilgaard 2004).

Several behavioural models in rats, which are predictive of antidepressant and anxiolytic activity, have shown similar results when ESC was administered with twice the concentration of R-CIT (Sánchez 2003).

Similarly, experiments using validated chronic models of antidepressant activity with high predictive value have shown similar effects and also indicate an earlier onset of action of ESC when compared to CIT (Montgomery *et al.* 2001).

The current evidence suggests that this inhibiting effect of R-CIT on ESC is exerted at the SERT itself. This has been established from several experiments using microdialysis techniques which demonstrated that R-CIT exerts its inhibitory effects in frontal cortex neuronal terminals. The data showed only a 300% rise above the baseline level of extracellular 5HT with R-CIT + ESC but a 450% rise with ESC alone, and receptor binding kinetic studies for (^{3}H)-ESC binding to the transporter (Mørk *et al.* 2003b; Wiborg & Sánchez 2003), in which the association rate of ESC was determined in the presence of increasing concentrations of R-CIT.

Although radioligand binding assays show ESC and R-CIT to have measurable affinity at the transporter binding sites, there is a marked difference: Ki for ESC was found to be 1.1 nM compared with 36 nM for R-CIT (Owens *et al.* 2001), indicating that R-citalopram has less affinity for the serotonin transporter. Similar results from in vivo binding studies suggest that transporter occupancy is high for ESC but not for R-CIT (~80% v ~30%) (Sánchez *et al.* 2004).

Several studies have shown that there are two binding sites for SSRIs on the SERT (Plenge & Mellerup 1985; Wennogle & Meyerson 1985; Plenge *et al.* 1991). One is a high affinity site, which is the primary binding site modulating transporter activity. The other is an allosteric site of low affinity for SSRIs that modulates the binding of drugs to the primary site. Binding of R-CIT to the allosteric site may bring about a conformational change in the transporter protein, which reduces the on-rate of ESC binding to the primary site (Wiborg & Sánchez 2003). This effect of ESC increasing inhibitor binding while citalopram reduces it appears to be more pronounced than with some other SSRIs (i.e. ESC strengthens its own binding). In experiments using (^{3}H)-ESC, (^{3}H)-paroxetine, (^{3}H)-fluoxetine, binding of the radioligand to human-SERT cell membranes and testing its dissociation rate from SERT when added to unlabelled test drug, the rate of dissociation of fluoxetine was rapid, while that of paroxetine was approximately half that of escitalopram (Chen *et al.* 2005). So, although more experiments are required to elucidate fully the mechanisms involved with the observed results concerning the effects of R- and S-citalopram at the target site, it is reasonably clear that the R-enantiomer has a counteractive effect against ESC. The problem is whether these animal data, showing an inhibitory effect of the R-isomer on the activity of the S-isomer, translate into greater efficacy of the latter compared to the racemate in human depression.

Clinical Studies – Review of Pooled Analyses

The animal data to support ESC as being an improvement on CIT is strong and, furthermore, suggests further possibilities for future research to find a fast-onset antidepressant with an improved side-effect profile. However, animal data does not necessarily translate into clinical practice. So does the animal data for ESC follow through to human clinical efficacy?

This study reviews five pooled analyses. Three groups of authors who performed systematic reviews of the clinical data available drew contradictory inferences from the same data (Gorman *et al.* 2002; Auquier *et al.* 2003; Svensson & Mansfield 2004). The trials used in these analyses and those used in the Gorman 2002 pooled analysis are listed in Table 15, which also shows the affiliation of the authors for each trial to Lundbeck. Auquier describes escitalopram as being an 'effective therapeutic treatment for MDD, presenting significant advantages over citalopram'. Surprisingly, he makes reference to two trials, one by Wade *et al.* (2002b) and one from Montgomery *et al.* (2001) but does not

include them in his analysis. The omission of the Wade trial results from the criterion that Auquier set to exclude trials with no active comparator. Although it is not explicit in the Auquier paper for the reason why the Montgomery data were omitted from his analysis, that data is only for the first four weeks of an 8-week trial and Auquier's criteria demanded a minimum of 8 weeks. A sensitivity analysis by including these omitted studies would have been useful to determine their effect on the result. It is perhaps of note that this paper was sponsored by Lundbeck.

Table 15: Trials included in published systematic overviews of escitalopram.

	Systematic Overview				
Trials	**Gorman[1] 2002**	**Auquier[2] 2003**	**Svensson[3] 2004**	**Lepola[1] 2004**	**Llorca[1] 2005**
Burke 2002	•	•	•	•	•
Colonna 2002		•	•		
Gorman 2002			•		
Lepola 2003		•		•	•
Montgomery 2001	•		•		
Wade 2002	•		•		
MD-02			•		
Data on File		•			•?

1 Industry affiliation 2 Unrestricted grant from Lundbeck 3 Independent

Conversely Svensson and Mansfield, in an independent study, consider the claims for ESC to be 'unwarranted' and they agree with the Swedish and Danish drug regulatory authorities that the clinical value of ESC and CIT appears to be equal and that there is no clear advantage of escitalopram over citalopram. This study by Svensson and Mansfield was a qualitative review of the studies available in 2003/4 and it focussed on the evidence from these trials in relation to the advertising claims being made. The authors particularly criticise the pooled analysis by Gorman *et al.* (2002) and the trial by Colonna *et al.* (2002). No main outcome is described in the Gorman paper; apart from stating that the main outcome in the trials they used for the analysis was mean change in MADRS score from baseline at week 8. As the Colonna paper was presented as a poster, it has a lack of sufficient detail for further analysis. Svensson included more trial data in his analysis,

including results from a failed trial that Lundbeck have not published (MD-02), although he did not include the Lepola data cited in Auquier. On inspection, it is probable that the 'unpublished data' quoted in the Auquier paper is the same as the MD-02 failed study that Svensson discusses. There are some similarities between them. There are the same numbers of patients (368 quoted in MD-02, 368 (from unpublished data) in Auquier) and each is described as a flexible dose study. MD-02 did include patients who were over 65 years old, which may have been a confounder, while other trials presented for the Food and Drug Administration submission had patient populations that had an upper limit of 65 years. There was no significant difference between placebo and either of the active drugs for the primary outcome measure (MADRS, LOCF). However, there was statistical significance for other parameters (MADRS OC, HAMD, CGI-I) and there was a numerical trend for greater response with escitalopram and citalopram compared with placebo. It is also not clear why MD-02 resulted in no significant differences between either ESC or the active comparator (CIT) and placebo but perhaps there were methodological problems that are not clear. However, there were twice as many adverse dropouts in the escitalopram group compared with the citalopram and placebo groups. There also appeared to be a large placebo effect that could not be accounted for (FDA submission data from Lundbeck, 2001).

The Gorman *et al.* pooled analysis was a pivotal piece of evidence used in Lundbeck's early promotional material and was one of only two references that supported claims of the superiority of ESC over CIT. Changes were made in later advertising material that included five references. Perhaps it is of note that the Gorman paper was not published in a top-rated journal and two of the authors were Forest/Lundbeck employees. The description of the methods used and the trials included is poor. There is no mention of the methods employed for randomisation in the individual trials, nor for allocation concealment or blinding, as should ideally be reported in clinical trial reports and meta-analyses according to the guidelines described in the CONSORT statement (Altman 1996; 2001; Moher 2001) (Appendix 1). Other key issues are missed, including compliance assessment, adverse effects, and withdrawal rates. The mean change in MADRS score from baseline to endpoint was the primary efficacy outcome in all three trials. The main outcome for the pooled analysis was not stated but presumably is mean change in MADRS score from baseline to endpoint. Intention to treat analysis was used for all patients receiving at least one dose of double-blind treatment with at least one post-baseline

MADRS assessment, using last observation carried forward. No data is reported for each individual trial but one is the failed MD-02 from the USA.

Gorman contends that ESC was statistically significantly superior to CIT in improving MADRS at week 1 in LOCF and also in week 6. However, although statistically significant, it is arguable that a one point difference between ESC and placebo is important clinically (Montgomery 1994). At week 6, the analysis by Gorman shows ESC having statistical superiority over placebo and CIT, and claims a trend in favour of ESC at this time point. However, he then fails to point out that at week 8, there is no longer statistical significance for ESC over CIT. Although the advantage over placebo remains at week 8, the MADRS score for ESC is only 0.8 points lower when compared to week 6 (-13.0 at week 6 vs. –13.8 ESC at week 8) and CIT is beginning to show convergence with escitalopram at week 8 (-13.8 ESC vs. –13.1 CIT). Table 16 shows the mean changes from baseline total score for the three treatment groups.

Table 16 Reduction in MADRS score relative to baseline for the three treatment groups

Study week	Placebo	Escitalopram	Citalopram
1	-3.8	-4.7	-3.7
2	-6.6	-7.8	-7.2
4	-9.4	-11.0	-10.2
6	-10.3	-13.0	-12.0
8	-11.2	-13.8	-13.1

Adapted from Gorman *et al.* 2002

Gorman refers many times to 'trends' in favour of ESC vs. CIT but then concludes that the result of his pooled analysis 'clearly supports' previous evidence of the antidepressant effect of ESC. There is an antidepressant effect, but probably to no greater extent than CIT. Apart from P values, no other statistical information (e.g. standard deviation, standard error or confidence intervals) is given so that further analysis is not possible.

Evaluation of trials and data abstraction

The literature search for this study identified thirteen clinical trials comparing ESC against various antidepressants and placebo, or assessing the effect of switching from another antidepressant to ESC, in a total population of 4833 subjects. Table 17 details the characteristics of the escitalopram trials. The 13 trials were reported in eight publications and seven posters; three of the posters were subsequently published in full. The MD-02

trial, a summary of which was downloaded from the FDA website, was described as a failed trial as the primary outcome measure (LOCF ITT dataset) and did not show statistically significant separation of either active drug (escitalopram or citalopram) from placebo. There was a numerical trend in favour of the two active compounds, but when compared to the citalopram and placebo arms there were also twice as many dropouts due to adverse events in the escitalopram group.

Trials Assessment

Each of the seven randomised controlled trials were inspected for methods, results and discussion using the criteria suggested by the CONSORT group as being indicative of proper trial methods and reporting (Moher *et al.* 2001). The results are described in Table 18; Appendix 3 contains the CONSORT data forms used to assess the papers. The methods sections of the papers were lacking in detail for the majority, the exception being the study by Moore *et al.* (2005). In fact, this paper only had three items missing: allocation concealment, implementation, and recruitment. They also included a clear flow diagram for patient disposition. The majority of studies did not include information about the randomisation or blinding processes, which makes it difficult to consider whether the conduct of the trial was correctly performed. (It would be assumed to be true.) This has been a common problem with trial reporting and the reason for the development of the CONSORT criteria (Altman 1996). Results were described generally well, although the recruitment of patients was not stated by any author. All described or mentioned ancillary analyses with the exception of the earlier Montgomery paper, but this was a preliminary report from four weeks' data of an eight-week trial. Discussions are rather variable in quality. All interpret the data but only three discuss the generalisability of the results and only four discuss the results in the context of current evidence.

Comparator Drugs

Citalopram is the main comparator (Burke *et al.* 2002; Colonna *et al.* 2002; Lepola *et al.* 2003; Montgomery *et al.* 2001; Moore *et al.* 2005; Reines *et al.* 2002). Reines and co-researcher Despiegel also presented work from a switch trial at the 3rd International Forum on Mood and Anxiety Disorders (Monte Carlo) in 2002 but this appears not to have been published: it was only available as a poster hand-out from the company. The Colonna study has since been published (Colonna *et al.* 2005), but the data used here are those from

the original poster. (These were data for the moderately ill group. These were used for the original promotional material, including the pooled analyses.) Reines *et al.* is a short report in European Neuropsychopharmacology and, although there are some data regarding numbers of patients in each arm and the doses used for the active drugs, there is insufficient information for useful analysis. Certainly there are no data to include the trial in the pooled analysis in this current study.

Citalopram, fluoxetine, paroxetine and sertraline were used in a cross-over study (Zimbroff *et al.* 2004), sertraline and venlafaxine used as direct comparators in three other trials (Alexopoulos *et al.* 2004; Bielski *et al.* 2004 and Montgomery *et al.* 2004). One trial (Rappaport *et al.* 2004) used an open-label 8-week phase of ESC before randomising in a double-blind trial to ESC or placebo for 36 weeks. Wade *et al.* (2002b) conducted a trial in primary care against placebo and also performed a continuation study over 52 weeks (Wade *et al.* 2002a). Of these trials, 10 are double-blind (Alexopoulos 2004; Bielski 2004; Burke 2002; Colonna 2002; Lepola 2003; Montgomery 2001, 2004; Moore 2005; Reines 2002 *et al.*; Wade 2002b), 5 are in outpatient settings (Bileski 2004; Burke 2002; Moore 2005; Rapaport 2004; Zimbroff 2004), 7 are in primary care (Colonna 2002; Lepola 2003; Montgomery 2001, 2004; Reines 2002 *et al.*; Wade 2002a, 2002b), one does not state which setting was used (Alexopoulos *et al.* 2004). The first Montgomery paper (2001) is a preliminary report prior to the publication of the full trial (Lepola *et al.* 2003).

Demographics & Inclusion/Exclusion Critria

The demographics of the study populations show that the duration of depression was only clearly defined in the studies by Alexopoulos *et al.* 2004 and Bielski *et al.* 2004,

Table 17a: Randomised controlled trials of escitalopram

	Author	Year	Setting	Design	Duration (wks)	Plac	Comp	Double Blind	Esc Dose	Comp Dose	N Esc/Comp/Plac (ITT)	Paper
1	Alexopoulos	2004	NS	PG	8		SER	•	10mg	50-200mg	104/108	
2	Bielski	2004	OP	PG	8		VNF	•	20mg	225mg	97/98	•
3	Burke (10mg)	2002	OP	PG	8	•	CIT	•	10mg	40mg	118/125/119	•
	Burke (20mg)	2002	OP	PG	8	•	CIT	•	20mg	40mg	123/125/119	•
4	Colonna	2002	PC	PG	24		CIT	•	10mg	20mg	85/85	
5	Lepola	2003	PC	PG	8	•	CIT	•	10-20mg	20-40mg	155/159/154	•
6	Montgomery	2001	PC	PG	4	•	CIT	•	10mg	20mg	155/159/154	•
7	Montgomery	2004	PC	PG	8		VNF	•	10-20mg	75-150mg	146/142	•
8	Moore	2005	OP	PG	8		CIT	•	20mg	40mg	138/142	•
9	Rapaport	2004	OP	PG	8			OL	10-20mg		502	
10	Reines	2002	PC	PG	8	•	CIT	•	10mg	20mg	155/160/154	
11	Wade	2002b	PC	PG	8	•		•	10mg		191/None/189	•

Table 17b: Cross-over or open label trials

9	Rapaport (extn)	2004	OP	PG	36	•		•	10-20mg		181/None/93	•
10	Reines (extn)	2002	PC	XO	8				10-20mg	•	104	
12	Wade	2002a	PC	XO	52				10-20mg		590	
13	Zimbroff (lead)	2004	OP	PG	8		CIT	OL		20-60mg	131	
	Zimbroff (lead)	2004	OP	PG	8		FLX	OL		20-80mg	129	
	Zimbroff (lead)	2004	OP	PG	8		PAR	OL		20-50mg	128	
	Zimbroff (lead)	2004	OP	PG	8		SER	OL		50-200mg	127	
	Zimbroff (extn)	2004	OP	XO	8			OL	10-20mg		136	

PG = Parallel Groups, XO = Cross-over, OL = Open Label
Comp = Comparator, Esc = Escitalopram, Pla = Placebo
NS= Not Specified, OP= Out Patients, PC= Primary Care

CIT = Citalopram, FLX = Fluoxetine, PAR = Paroxetine
SRT = Sertraline, VNF = Venlafaxine

Table 18: CONSORT Analysis for Escitalopram RCTs									
Paper Section and topic		Item	Bielski 2004	Burke 2002	Lepola 2003	Montgomery 2001	Montgomery 2004	Moore 2005	Wade 2002b
Title & Abstract		1	•	•	•	•	•	•	•
Introduction (Background)		2	•	•	•	•	•	•	•
Methods									
	Participants	3	•	•	•	•	•	•	•
	Interventions	4	•	•	•	•	•	•	•
	Objectives	5	•	•	ns	ns	ns	•	ns
	Outcomes	6	•	•	•	•	•	•	•
	Sample Size	7	ns	ns	•	•	•	•	ns
	Randomization	8	ns	ns	ns	ns	ns	•	ns
	Allocation concealment	9	ns	ns	ns	ns	ns	ns	ns
	Implementation	10	ns	ns	ns	ns	ns	ns	ns
	Blinding	11	•	ns	ns	ns	•	•	ns
	Statistics	12	•	•	•	•	•	•	•
Results									
	Participant flow	13	•	•	•	•	•	•	•
	Recruitment	14	ns	ns	ns	ns	ns	ns	ns
	Baseline data	15	•	•	•	ns	ns	•	•
	Numbers analysed	16	•	•	•	•	•	•	•
	Outcomes & estimation	17	•	•	•	•	•	•	•
	Ancillary analyses	18	•	•	•	ns	•	•	•
	Adverse events	19	•	•	•	•	•	•	•
Discussion									
	Interpretation	20	•	•	•	•	•	•	•
	Generalizability	21	ns	•	ns	ns	ns	•	•
	Overall evidence	22	ns	•	ns	ns	•	•	•
	ns = not specified								

Table 19a: Summary of baseline patient characteristics

Author	Duration of depression (years)			% Recurrent			% Female			Mean age			Mean Baseline MADRS		
	ESC	Comp	PBO	ESC	Comp	PBO	ESC	Comp	PBO	ESC	Comp	PBO	ESC	Comp	PBO
Alexopoulos	8.9	7.4	0.0	52.9	47.2	0.0	54.8	60.2	0.0	40.6	38.1	0.0	29.5	29.0	0.0
Bielski	8.5	9.8	0.0	60.2	62.0	0.0	69.4	47.0	0.0	37.3	37.5	0.0	30.7	30.0	0.0
Burke	>4 week duration			71.0	70.0	69.0	69.0	62.0	60.0	40.1	40.0	40.1	28.5	29.2	29.5
Colonna	——— ns ———			——— ns ———			73.0	76.0	0.0	46.0	46.0	0.0	26.3	26.3	0.0
Lepola	——— ns ———			——— ns ———			74.8	69.4	72.1	43.0	44.0	43.0	29.0	29.2	28.7
Montgomery '01	——— ns ———			——— ns ———			30.0	30.0	30.0	43.0	43.0	43.0	29.0	29.0	29.0
Montgomery '04	——— ns ———			——— ns ———			73.0	71.0	0.0	49.0	47.0	0.0	28.7	29.0	0.0
Moore	——— ns ———			45.7	44.4	0.0	71.7	62.0	0.0	44.1	46.2	0.0	36.3	35.7	0.0
Rapaport	>4 week duration			70.0	0.0	68.0	60.2	0.0	62.4	42.9	0.0	41.8	7.2	0.0	6.2
Reines	——— ns ———			——— ns ———			——— ns ———			——— ns ———			29.0	29.0	29.0
Reines extn	——— ns ———			——— ns ———			77.0	73.0	0.0	43.0	46.0	0.0	10.4	11.8	0.0
Wade 2002a	——— ns ———			——— ns ———			75.0	0.0	0.0	42.0	0.0	0.0	14.2	0.0	0.0
Wade 2002b	——— ns ———			——— ns ———			73.8	0.0	77.8	41.0	0.0	40.0	29.2	0.0	28.7
Zimbroff	——— ns ———			——— ns ———			65.0	0.0	0.0	41.0	0.0	0.0	22.2	0.0	0.0
	Average			60.0	44.7	27.4	*69.7*	*65.1*	*60.5*	*42.5*	*43.1*	*41.6*	*24.7*	*27.7*	*24.4*
	Average/grp				44.0			*65.1*			*42.4*			*25.6*	

PBO = placebo Comp = comparator ns = not specified

Table 19b: Inclusion and exclusion criteria and outcomes for Escitalopram RCTs

Author	Inclusion Criteria	Exclusion Criteria	Outcomes Measured				
			HAMD	MADRS	Responders	Remitters	Other
Alexopoulos	OPs 18-80yrs; DSM-IV MDD MADRS ≥22	ns	Secondary measure	Total score LOCF at wk 8	>50% decrease from baseline	MADRS total score <12	ns
Bielski	OPs 18-65yrs; DSM-IV MDD HAMD(24)≥20	Abnormal lab tests, pregnancy. Other psychiatric disorder	Secondary measure	Total score LOCF at wk 8	>50% decrease from baseline	MADRS total score <12	CGI, HAMA, CES-D, Q-LES-Q
Burke	OPs 18-65yrs; DSM-IV MDD MADRS≥22 min score of 2 on HAMD item1	Abnormal lab tests, pregnancy. Other psychiatric disorder. Other psychotropic meds. Suicidal ideas.	Secondary measure	Total score LOCF at wk 8	>50% decrease from baseline	MADRS total score <12	CGI, HAM-A, CES-D, QOL
Colonna	Primary care 18-65yrs; DSM-IV MDD MADRS ≥22	ns	x	Total score LOCF at wk 24	>50% decrease from baseline	MADRS total score <12	CGI, HAM-A
Lepola	Primary care 18-65yrs; DSM-IV MDD MADRS≥22 ≥40	Other psychiatric disorder. Suicidal ideas.	x	Total score LOCF at wk 8	>50% decrease from baseline	MADRS total score <12	CGI
Montgomery 2001	Primary care; DSM-IV MDD MADRS ≥22 ≥40	ns	x	Total score LOCF at wk 4	ns	ns	CGI

Table 19b: Inclusion and exclusion criteria and outcomes for Escitalopram RCTs (contd)

Montgomery 2004	Primary care 18-65yrs; DSM-IV MDD MADRS ≥18	Other psychiatric disorder. Suicidal ideas. Other psychotropics	Secondary measure	Total score LOCF at wk 8	>50% decrease from baseline	MADRS total score <12	none
Moore	OPs & Primary care 18-65yrs; DSM-IV MDD MADRS ≥30. Physically well.	Other psychiatric disorder. Other psychotropics. Substance abuse.	x	Total score LOCF at wk 8	>50% decrease from baseline	MADRS total score <12	CGI-S, MADRS-S
Reines	Primary care; DSM-IV MDD MADRS ≥22 ≥40	ns	x	Total score LOCF at wk 8	>50% decrease from baseline	MADRS total score <12	CGI
Wade 2002b	Primary care 18-65yrs; DSM-IV MDD MADRS ≥22 ≥40 Physically well	Other psychiatric disorder. Suicidal ideas. Other psychotropics	x	Total score LOCF at wk 8	>50% decrease from baseline	MADRS total score <12	CGI-S

ns = not specified
x = not used
MDD = major depressive disorder
MADRS = Montgomery-Åsberg depression rating scale
MADRS-S = MADRS Self-rating
HAMD = Hamilton depression rating scale
HAMA = Hamilton anxiety rating scale
OPs = outpatients
LOCF = last observation carried forward

CGI-I = Clinical Global Impression- Improvement
CGI-S = Clinical Global Improvement – Severity
CES-D = Centre for Epidemiological Studies-Depression Scale
Q-LES-Q = Quality of Life Enjoyment and Satisfaction Questionnaire
QOL = Quality of Life questionnaire

Table 20a: MADRS Baseline and End-point Scores (versus Comparator)

Author	Year	Comparator	Escitalopram						Comparator					
			Baseline	SD	End Point Reduction	SD	End Point actual	Sample Size	Baseline	SD	End Point Reduction	SD	End Point actual	Sample Size
Alexopoulos	2004	Sertraline	29.5	±4.3	-16.9	ns	12.6	104	29.0	±4.0	-16.1	ns	12.9	107
Bielski	2004	Venlafaxine	30.7	±4.6	-15.9	±10.3	14.8	98	30.0	±5.0	-13.6	±9.6	16.4	100
Burke (10mg)	2002	Citalopram	28.0	±4.9	-12.8	±0.8	15.2	118	29.2	±4.5	-12.0	±0.9	17.2	125
Burke (20mg)	2002	Citalopram	28.9	±4.6	-13.9	±0.8	15	123	29.2	±5.0	-12.0	±0.9	17.2	125
Colonna	2002	Citalopram	26.3	ns	-21.6	ns	4.7	165	26.3	ns	-20.6	ns	5.7	174
Lepola	2003	Citalopram	29.0	ns	-15.0	ns	14	155	29.2	ns	-13.6	ns	15.6	160
Montgomery	2001	Citalopram	29.0	ns	-11.3	ns	17.7	155	29.0	ns	-10.2	ns	18.8	160
Montgomery	2004	Venlafaxine	28.7	±5.0	-18.2	ns	10.5	146	29.0	±5.4	-18.9	ns	10.1	145
Moore	2005	Citalopram	36.3	±4.8	-22.4	ns	13.9	138	35.7	±4.4	-20.3	ns	15.4	142
Rapaport (open)	2004	Citalopram	13.8	±8.6	-3.2	±8.5	10.6	219	14.0	±10.2	-2.9	±8.6	11.1	138
Reines (extn)	2002b	Citalopram	10.4	ns	-3.8	ns	6.6	52	11.8	ns	-5.6	ns	6.2	52
		Mean	*29.6*		*-16.4*		*13.2*		*29.6*		*-15.3*		*14.4*	
		SD	*2.8*		*3.8*		*3.7*		*2.5*		*3.9*		*4.2*	

ns = not specified SDs as reported in the original report

The means do not include the open or extension studies of Rapaport, Reines

Table 20b: MADRS Baseline and End-point Scores (versus Placebo)

Author	Year	Comparator	Escitalopram						Placebo					
			Baseline	SD	End Point Reduction	SD	End Point actual	Sample Size	Baseline	SD	End Point Reduction	SD	End Point actual	Sample Size
Burke (10mg)	2002	Citalopram	28.0	±4.9	-12.8	±0.8	15.2	118	29.5	±5.0	-9.4	±0.9	20.1	119
Burke (20mg)	2002	Citalopram	28.9	±4.6	-13.9	±0.8	15	123	29.5	±5.0	-9.4	±0.9	20.1	119
Lepola	2003	Citalopram	29.0	ns	-15.0	ns	14	155	28.7	ns	-12.1	ns	16.6	154
Montgomery	2001	Citalopram	29.0	ns	-11.3	ns	17.66	155	29.0	ns	-8.7	ns	20.32	154
Rapaport (open)	2004	Citalopram	13.8	±8.6	-3.2	±8.5	10.6	219	17.4	±9.7	-5.5	±10.9	11.9	145
Rapaport (extn)	2004	Placebo	7.2	±4.0	-1.4	±4.8	5.8	181	6.2	±3.8	0.7	±4.2	6.9	93
Wade	2002b	Placebo	29.2	ns	-14.9	ns	14.3	191	28.7	ns	-12.0	ns	16.7	189
		Mean	*26.3*		*-11.9*		*14.5*		*27.1*		*-9.5*		*17.6*	
		SD	*6.1*		*4.5*		*2.3*		*4.8*		*2.4*		*3.3*	

na = Not Applicable

The means do not include the open or extension studies of Rapaport

while those by Burke *et al.* 2002 and Rapaport *et al.* 2004 report that patients entered into their trials had depressive episodes for longer than four weeks. The remaining studies have no information on the duration of illness (Table 19a). The lack of information here does not give any indication of the severity of the illness in terms of time, which otherwise might indicate a severe, longer-lasting illness.

Table 19b summarises the inclusion and exclusion criteria, and outcomes measured in the ten randomised controlled trials. It can be seen that all use similar inclusion criteria, although the age ranges differ slightly. The gender difference was similar across the studies averaging 65% female. The exception was the Montgomery *et al.* study of 2001, as the data was averaged out across the three treatment arms at 30%. The mean age and the mean baseline MADRS score are approximately 42 years and 25.5 points, respectively. Montgomery 2001 does not state the ages of participants, while the Alexopoulos study could include subjects up to the age of eighty years old. However this did not alter the mean age when compared to those in the other studies. The posters and short reports did not state any exclusion criteria. The published papers did and showed similarity in the exclusion criteria used. All used MADRS as the primary outcome measure, four employing the Hamilton rating scale as a secondary measure.

Recurrence of illness is only reported in five studies (Alexopoulos *et al.* 2004; Bielski *et al.* 2004; Burke *et al.* 2002; Moore *et al.* 2005; Rapaport *et al.* 2004). In four studies, the percentage recurrence was comparable between groups within each study, except for the trial by Alexopoulos *et al.* where there was a 5.7% difference between the escitalopram and sertraline groups.

Efficacy

Table 20a describes the efficacy data in terms of the MADRS end-point scores for escitalopram versus comparator; Table 20b summarises the same data for escitalopram versus placebo. The standard randomised controlled trials comparing escitalopram against an active comparator, placebo, or both, have comparable baseline MADRS scores. End-point scores are based on last observation carried forward (LOCF), intention-to-treat (ITT). The reductions in these scores at end-point are very similar across these studies. Also the standard deviations (which are those in the reports) are very similar numerically.

There are small differences in MADRS total scores between escitalopram and comparator, except for the cohort that was given 20mg escitalopram in Burke *et al.*'s study, where there is a 2.2 points difference (Table 21). There is better separation between

escitalopram and placebo in the Burke study at both doses, but the separation is lower in the other trials. An effect size of 3-4 points is generally considered as clinically relevant, although the size of effect may be reduced in some circumstances: eg higher than expected numbers of drop-outs, particularly early in treatment (Montgomery 1994). If the study population included an unusually high proportion of resistant depression, this might also reduce the effect size. The difference in effect size between active and placebo may also reduce if the placebo response was unusually large. However, these studies do not appear to have large numbers of drop-outs. Table 22 shows the numbers of overall drop-outs in each of the studies where these have been reported. Two studies reported withdrawals due to adverse effects specifically (Burke *et al.* 2002; Zimbroff 2004). The 153 subjects in the Wade *et al.* study (2002a) represent 26% of the study population.

Table 21:
MADRS End-point Differences between ESC and either comparator or placebo

Author	Year	Comparator	Difference Comparator	PBO
Alexopoulos	2004	Sertraline	-0.3	na
Bielski	2004	Venlafaxine	-1.6	na
Burke (10mg)	2002	Citalopram	-2	-4.9
Burke (20mg)	2002	Citalopram	-2.2	-5.1
Colonna	2002	Citalopram	-1	na
Lepola	2003	Citalopram	-1.6	-2.6
Montgomery	2001	Citalopram	-1.13	-2.66
Montgomery	2004	Venlafaxine	0.47	na
Moore	2005	Citalopram	-1.5	na
Rapaport (open)	2004	Citalopram	-0.5	-1.3
Rapaport (extn)	2004	Placebo	na	-1.1
Reines et al	2002	Citalopram	1.5	2.9
Reines (extn)	2002	Citalopram	0.4	na
Wade (extn)	2002a	Open-label	na	na
Wade	2002b	Placebo	na	-2.4

na = Not Applicable
A minus sign favours escitalopram

Lepola *et al.* (2003) reported the full eight week analysis that included data from an earlier four-week study reported by Montgomery *et al.* (2001), which was a preliminary mid-study analysis of the 8-week trial. The Montgomery paper compared the results from animal data using the chronic mild stress model of depression in rats, which showed a rapid onset of action, with early human trial data that appeared to show a similar rapid onset. Lepola found statistically significant superiority of escitalopram versus citalopram (as well as placebo) in the responder analysis. However, the study was not powered to

Table 22: Time to Response and Number of Drop-outs

Author	Year	Comparator	Response Time (days)			Number drop-outs			
			ESC	Comp	PBO	ESC	Comp	PBO	
Alexopoulos	2004	Sertraline	24.5	31.5	na	16	15	na	
Bielski	2004	Venlafaxine	47.6	38.5	na	26	34	na	
Burke (10mg)	2002	Citalopram	0	0	0	5	11	3	*
Burke (20mg)	2002	Citalopram	0	0	0	13	-	-	
Colonna	2002	Citalopram	38.5	49	na	10	26	na	
Lepola	2003	Citalopram	44.8	53.2	0	9	8	15	
Montgomery	2001	Citalopram	0	0	0	0	0	0	
Montgomery	2004	Venlafaxine	28	28	0	0	0	0	
Moore	2005	Citalopram	0	0	0	6	15	0	
Rapaport (extn)	2004	Placebo	na	na	na	7	na	7	
Reines (extn)	2002b	Citalopram	na	na	na	0	0	na	
Wade (extn)	2002a	Open-label	na	na	na	153	na	na	
Wade	2002b	Placebo	29.4	0	>56	31	0	29	
Zimbroff (lead)	2004	Citalopram	0	0	0	na	43	na	*
Zimbroff (lead)	2004	Fluoxetine	0	0	0	na	21	na	*
Zimbroff (lead)	2004	Paroxetine	0	0	0	na	20	na	*
Zimbroff (lead)	2004	Sertraline	0	0	0	na	15	na	*
Zimbroff (extn)	2004		0	0	0	27	na	na	

na = Not Applicable
* due to adverse events

detect differences between the active compounds. Studies powered for head-to-head comparisons are needed to demonstrate this. A concern with the report of this study is the confusing way in which numerical results are used to support their arguments. Lepola *et al.* conclude that ESC shows statistical superiority to CIT and is well tolerated. They further suggest that ESC is appropriate first-line treatment in primary care and infer that it should replace CIT. However, as the latter is available as a generic with significantly lower acquisition costs, the efficacy argument becomes more difficult to sustain. Lepola *et al.* do not address this issue.

The latest study has been specifically designed to examine the efficacy and tolerability of ESC versus CIT in an outpatient setting (Moore *et al.* 2005). No placebo arm was used, but it was assumed that ESC has proven efficacy over placebo. In order to reduce any placebo effect or spontaneous remission, only patients with a MADRS score >30 were eligible for inclusion. It has been reported that the higher the MADRS cut-off value, the better the chance of patients responding to the active compound only (Montgomery 1999a). The authors estimated the sample size by using results from a previous study by Burke *et al.* (2002), which compared efficacy of ESC, CIT and placebo. Taking account of attrition due to withdrawal gave an estimated sample size of 280

patients. Once titrated to the fixed doses of 20mg escitalopram and 40mg citalopram after the first week, dose alteration was not allowed.

The change in MADRS total score from baseline to end-point using last observation carried forward was used as the primary outcome measure. Absolute values of MADRS at weeks 1, 4, 8 and LOCF, treatment response and remission were used as secondary measures. Other secondary outcomes included a self-rating version of MADRS, MADRS-s, and the change from baseline to end-of-study scores for Clinical Global Impression of Severity scale (Guy 1976).

Moore *et al.* found a higher percentage of patients completed the trial in the escitalopram group than in the citalopram group (95.7% versus 89.4%; P=0.047). However, although lack of efficacy leading to withdrawal was apparently four times more likely in the citalopram group, this difference was not statistically significant (P=0.19). Similarly, citalopram patients were not more significantly likely to withdraw due to adverse events (P=0.17).

Responder rates were high for escitalopram compared to citalopram (76.1% versus 61.5%) and highly significant (P=0.009), although for remitters the unadjusted rates for ESC versus CIT were 54.3% versus 43.0%, of borderline statistical significance (P=0.06). When using the adjusted initial MADRS values and physician specialisation (psychiatrist and GP), remittance attained statistical significance (56.1% for ESC versus 43.6% for CIT; P=0.04).

Switch Trials

There are three principal switch trials that were reviewed as part of this study. They extended an initial parallel group phase by a cross-over phase (Reines & Despiegel 2002; Wade *et al.* 2002a; Zimbroff *et al.* 2004). Zimbroff switched subjects from an open-label parallel group phase to an open-label variable dosage phase of ESC 10-20mg over 8 weeks. The Reines study investigated a switch from CIT, ESC or placebo after an eight-week double-blind lead-in. This trial was designed to demonstrate that patients could be switched from CIT to ESC with no deterioration in efficacy and with no change in tolerability. Although this was demonstrated, it is not clear that if patients had remained on citalopram, they would not have had similar outcomes.

The third switch trial (Wade *et al.* 2002a) used subjects from two European studies that had used either flexible dosed CIT or ESC versus placebo (Lepola *et al.* 2003), or used fixed-dose ESC versus placebo (Wade *et al.* 2002b). Both were conducted over an eight-

week period. The extension period was 52 weeks, using flexible-dose, open-label ESC. The baseline mean MADRS score for the extension study was 14.2, representing responders rather than remitters. Zimbroff *et al.* looked at non-responders with a mean MADRS score of 22.2 at the start of the switch period, while Reines *et al.* used responders who had a switch start point MADRS score of 11.8 for citalopram and 10.4 for escitalopram groups (OC values; LOCF may have been larger but not reported). 60% of patients in the Wade study were in remission by week 4 of the extension period, with the mean MADRS score of 12, defined in the trial as remission. It is not clear if patients achieving this had had a dose increase to 20mg during the four-week period, which may have improved the chance of remitting.

A fourth switch trial examined the efficacy and tolerability when switching from one of four SSRIs to ESC (Rosenthal & Li 2002). The data from this trial was presented at two different conferences, but this was perhaps not always made clear when the posters were given out as part of promotional material: at first sight, it looked like two new and separate trials. A small number of patients was recruited (46) who had discontinued one of citalopram, fluoxetine, paroxetine, or sertraline in an 8 week open-label trial. A patient discontinuing from one of these SSRIs due to adverse events was switched to open-label ESC. There almost appears to have been an assumption of at least same efficacy of escitalopram compared with the previous antidepressant. Of the 46 patients, 39 (85%) were successfully switched to ESC with no further adverse events that would have caused discontinuation. The focus does seem to have been on the lack of adverse events on switching. Depression symptoms improved during the ESC treatment period but this does not mean that ESC was necessarily any better than the former drug. Patients could well have improved on their former drug if side effects had not been a problem. No efficacy data was presented except for a graph of the mean MADRS scores. There is no numerical detail to perform further analysis and this trial does not seem to have been published.

Table 23: Responders & Remitters (%)

Author	Year	Comparator	Escitalopram		Comparator		Placebo	
			n(%) Responders	n(%) Remitters	n(%) Responders	n(%) Remitters	n(%) Responders	n(%) Remitters
Alexopoulos	2004	Sertraline	78(75)	63(61)	70(65.3)	64(59.2)	na	na
Bielski	2004	Venlafaxine	57(58.8)	40(50.5)	47(48)	36(41.8)	na	na
Burke (10mg)	2002	Citalopram	59(50)	ns	57(45.6)	ns	ns	ns
Burke (20mg)	2002	Citalopram	63(51.2)	ns	57(45.6)	ns	ns	ns
Colonna	2002	Citalopram	70(82)	70(82)	63(74)	60(71)	na	na
Lepola	2003	Citalopram	99(63.7)	81(52.1)	84(52.6)	68(42.8)	99(48.2)	ns
Montgomery	2004	Venlafaxine	113(77.4)	102(69.9)	113(79.6)	99(69.7)	na	na
Moore	2005	Citalopram	105(76.1)	75(56.1)	87(61.5)	61(43.6)	na	na
Wade	2002b	Placebo	103(55)	103(47.5)	na	na	42(42)	79(34)

na = Not Applicable
ns = Not Specified

Meta-analysis of the RCTs

For the purpose of this study, only eight trials of the twelve identified have sufficient information to analyse the remitter and responder data in the active groups (Table 23). The Bielski study has been split into two trial groups based on escitalopram 10mg and 20mg for this analysis. The cross-over studies have no comparator in the switched phase, except for the extension study of Rapaport, in which patients were blinded to receive either escitalopram or placebo. Therefore the current analysis only considers the RCTs and does not include the switch trials. The summary data in terms of percentages for responders and remitters are described in Table 23 for those trials that have the appropriate data to extract. Of the placebo controlled trials, only Lepola *et al.* (2003) and Wade *et al.* (2002b) have extractable data for the placebo arm. The percentages in this table are converted to numbers of patients in the 2x2 contingency tables (see Appendix 4). The primary outcome measure is the MADRS score for patients who responded or remitted.

There is a trend in favour of escitalopram for responders, particularly against placebo. A similar trend exists for remitters. However, Montgomery (2004) is less clear-cut, with responder and remitter rates being similar between the escitalopram and venlafaxine arms. The Zimbroff *et al.* study (2004) rates are not shown here as the study was designed as a switch from initial SSRI treatment to escitalopram only. In the period of escitalopram treatment after the switch from one of citalopram, fluoxetine, paroxetine or sertraline, 60% responded and 41% were in remission at the end of the 8-week period. There are only small numbers of patients in each arm (30, 42, 32, 32, respectively) giving a total of 136 patients in the open-label treatment phase. The responder rate for the total

population is comparable to those in other studies. However, the remitter rate of only 41% is lower than most remitter rates in the other studies with the exception of Wade *et al.* (2002b), which had a rate of 47.5%. Curiously, the group switched from sertraline had the highest response and remission rates (70% and 57%, respectively) compared to the other three groups. The overall response in this study seems to indicate that escitalopram is no better that the comparators, although it does show that switching to escitalopram would probably not cause a relapse of depressive illness. That continuation of active treatment helps to prevent relapse is also borne out by the Rapaport *et al.* study (2004), which randomised patients with remitted depression to placebo or escitalopram and measured the time to depression relapse (defined as a MADRS score $\geq$ 22). Those on active treatment not only had a reduced chance of relapse but also demonstrated a further, small, reduction in MADRS scores, inferring that continued treatment may also have benefits other than preventing relapse. Reines and Despiegel (2002) appear to confirm further reduction in MADRS scores in their open-label extension study.

Analysis of the responders data shows that a fixed-effects model is appropriate to describe the data. The low values for the Cochran Q statistic (5.8) and I^2 (0%) for the active comparator responders data show there is no heterogeneity between the studies, although inconsistency is greater for the placebo responders odds ratio (I^2 = 42%). For the seven studies that have sufficient data for analysis of escitalopram versus active comparator for responders (Table 24A), the pooled point estimate for the odds ratio (OR) is 1.41 (95% CI 1.17-1.71; P = 0.0004), indicating an approximately 40% greater chance of effect with escitalopram than with active comparator. Against placebo, the pooled estimate for the OR for the three trials in which placebo was used (Table 24B) is 2.05 (95% CI 1.61-2.61); i.e. escitalopram is twice as likely as placebo to produce an antidepressant effect. The Forest plot describing these data is shown in Fig 12b, which shows that all data sets lie well to the right of the null value (unity): escitalopram is better than placebo.

For the current study, the data from the only meta-analysis published (Auquier *et al.* 2003) were compared with the data derived here. Auquier *et al.* performed sensitivity analyses, including one that omitted the failed MD-02 trial. This particular analysis left the Burke *et al.* 2002, the Colonna *et al.* 2002, and the Lepola *et al.* 2003 trials, which were analysed as a subset of the seven trials. The odds ratios for response rate are 1.38 (95% CI = 1.06 to 1.79; P = 0.02) in the current study and 1.35 (95% CI = 1.09 to 1.70 P = 0.003) in the Auquier study. These compare with a pooled odds ratio for the eight trials (Burke being split into two, one for each escitalopram dose) of 1.41, which is in favour of escitalopram.

Table 24: Odds Ratios, Rate Ratios and Rate Differences, Responders

A: Active Comparator

Author	Year	Comparator	OR			RR			RD		
				Lower CI	Upper CI		Lower CI	Upper CI		Lower CI	Upper CI
Alexopoulos	2004	Sertraline	1.59	0.87	2.88	1.15	0.11	3.76	-0.10	-0.10	10.24
Bielski	2004	Venlafaxine	1.55	0.88	2.72	1.23	0.09	4.44	-0.11	-0.12	9.02
Burke (10mg)	2002	Citalopram	1.19	0.72	1.97	1.10	0.12	3.89	-0.04	-0.05	22.63
Burke (20mg)	2002	Citalopram	1.25	0.76	2.06	1.12	0.11	3.98	-0.06	-0.06	17.67
Colonna	2002	Citalopram	1.63	0.78	3.41	1.11	0.11	3.56	-0.08	-0.09	11.97
Lepola	2003	Citalopram	1.58	1.00	2.48	1.21	0.09	4.05	-0.11	-0.12	8.83
Montgomery	2004	Venlafaxine	0.88	0.50	1.54	0.97	0.15	2.98	0.02	0.02	-45.83
Moore	2005	Citalopram	2.01	1.20	3.37	1.24	0.09	4.07	-0.15	-0.15	6.45

B: Placebo

Author	Year	Comparator	OR			RR			RD		
				Lower CI	Upper CI		Lower CI	Upper CI		Lower CI	Upper CI
Burke (10mg)	2002	Citalopram	2.79	1.63	4.77	1.89	0.02	9.37	-0.24	-0.24	3.76
Burke (20mg)	2002	Citalopram	2.93	1.72	4.98	1.94	0.02	9.78	-0.25	-0.26	3.53
Lepola	2003	Citalopram	2.03	1.29	3.19	1.37	0.07	4.84	-0.17	-0.18	5.42
Wade (extn)	2002b	Open-label	4.24	2.71	6.63	2.47	0.01	15.83	-0.33	-0.33	2.42

OR=odds ratio. RR=rate ratio. RD=risk difference. CI=confidence interval

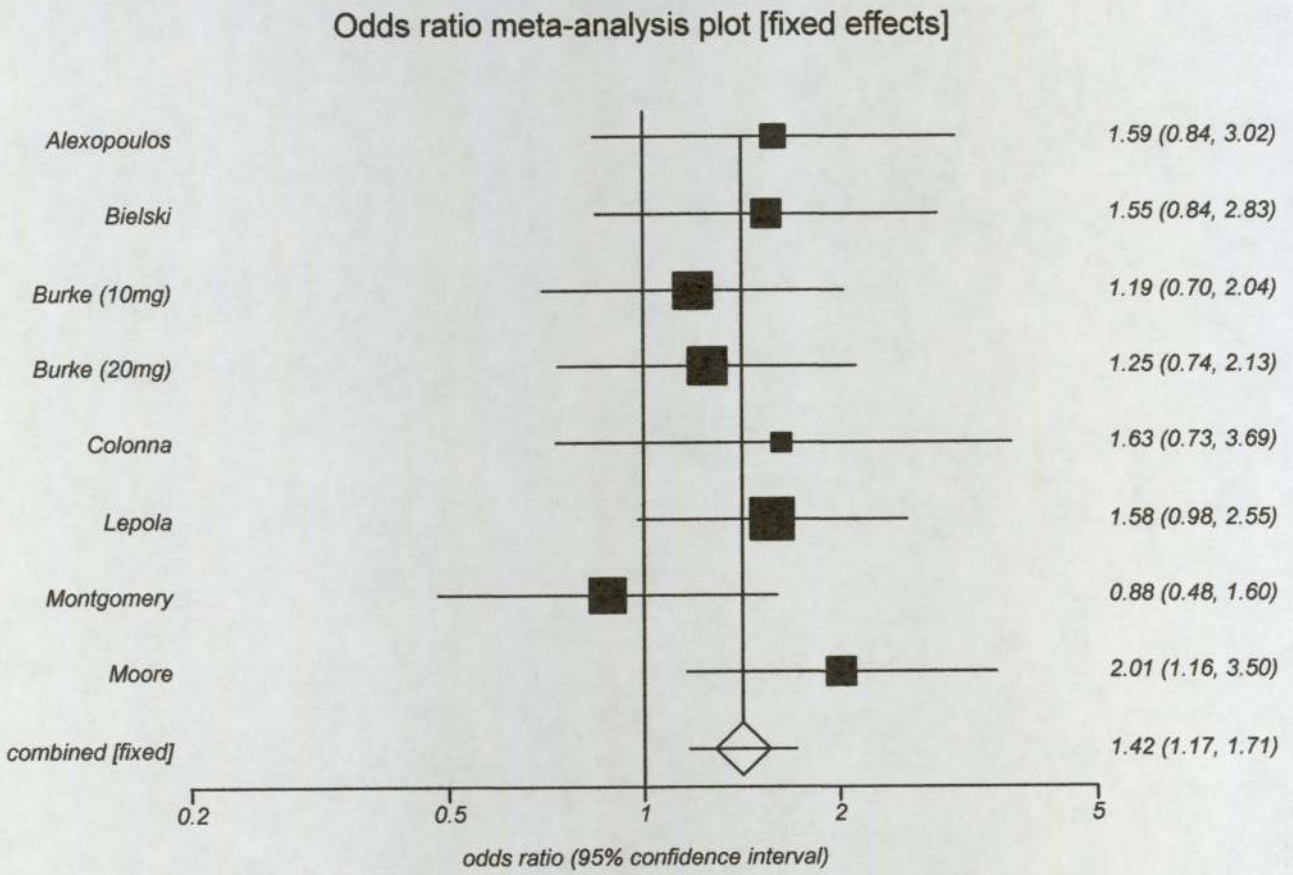


Fig 12a: Forest plot for escitalopram versus active comparator for the Responder groups

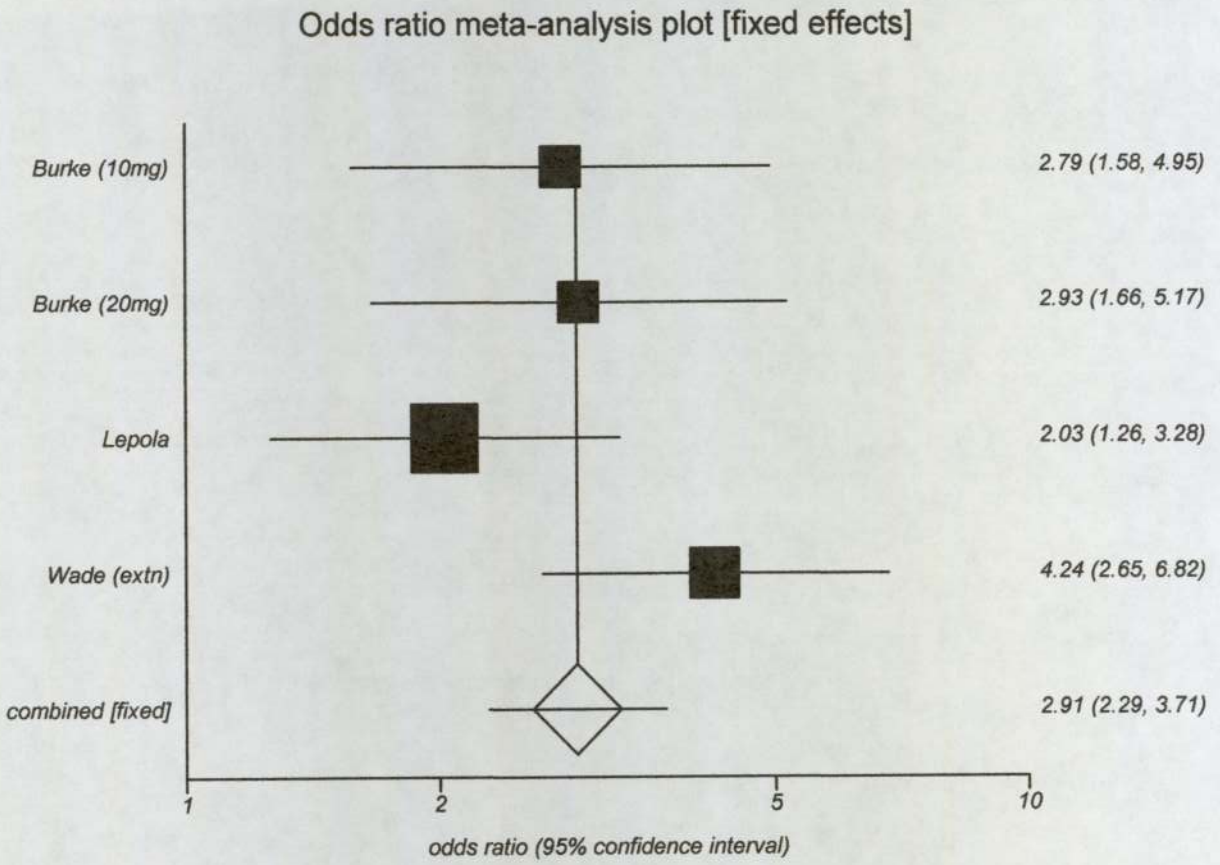


Fig 12b: Forest plot for escitalopram versus placebo for Responder groups

Table 25: Odds Ratios, Rate Ratios and Rate Differences, Remitters

A: Active Comparator											
Author	Year	Comparator	OR			RR			RD		
				Lower CI	Upper CI		Lower CI	Upper CI		Lower CI	Upper CI
Alexopoulos	2004	Sertraline	1.06	0.61	1.83	1.02	0.13	3.46	-0.01	-0.02	75.85
Bielski	2004	Venlafaxine	1.21	0.68	2.15	1.12	0.11	4.37	-0.05	-0.05	22.10
Colonna	2002	Citalopram	1.94	0.94	4.02	1.17	0.10	3.80	-0.12	-0.13	8.25
Lepola	2003	Citalopram	1.46	0.94	2.29	1.22	0.09	4.29	-0.09	-0.10	10.34
Montgomery	2004	Venlafaxine	1.01	0.61	1.67	1.00	0.14	3.17	0.00	-0.01	691.05
Moore	2005	Citalopram	1.58	0.99	2.53	1.27	0.08	4.52	-0.11	-0.12	8.54
B: Placebo											
Author	Year	Comparator	OR			RR			RD		
				Lower CI	Upper CI		Lower CI	Upper CI		Lower CI	Upper CI
Wade (extn)	2002b	Open-label	1.69	1.12	2.54	1.31	0.08	4.59	-0.13	-0.13	7.43

OR=odds ratio. RR=rate ratio. RD=risk difference. CI=confidence interval

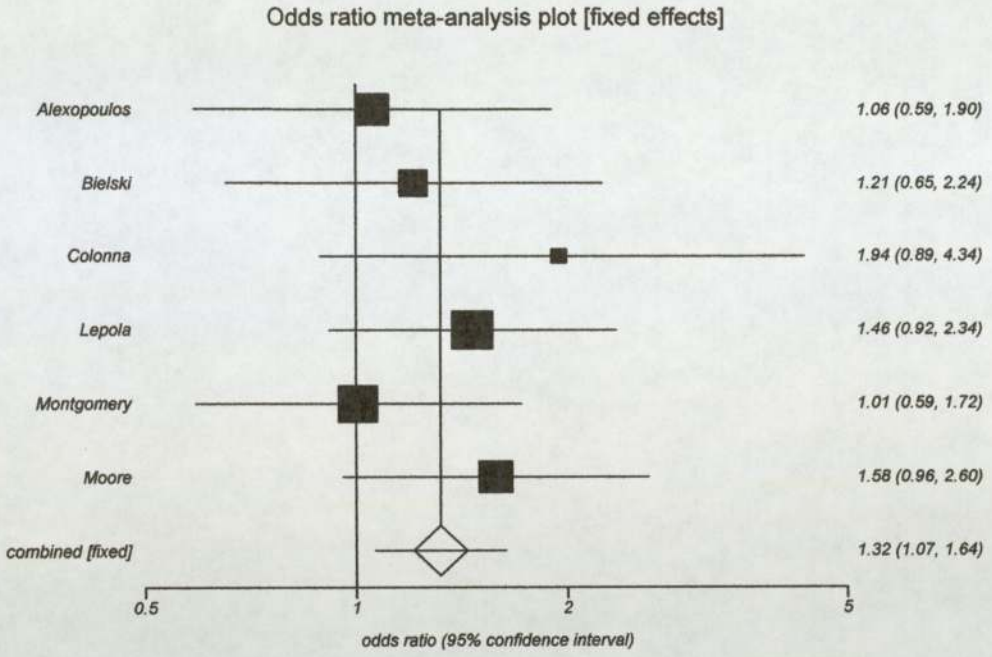


Fig 13: Forest plot for Remitters (active comparator)

As seen from Figure 12a, apart from the Moore *et al.* trial (2005), the majority of studies, although positive for escitalopram, have confidence intervals that pass through 1, indicating non-significance. However the combined effect shows a positive effect with a narrow confidence interval that does not pass through 1. Figure 12b shows that escitalopram is significantly better than placebo.

Remission

When considering remission, only the data for active comparators can be used for a pooled analysis, as there is only one placebo-controlled trial for remission, which is the Wade *et al.* (2002a) open label extension study. The odds ratio for this current study is 1.69 (95% CI = 1.12 to 2.54). The pooled odds ratio for remitters receiving active comparators is 1.32 (95% CI = 1.07 to 1.64; P = 0.012). The data are described in the Forest plot, Figure 13. This shows that two of the studies have odds ratios close to 1, inferring that escitalopram has no greater efficacy than the comparators (sertraline in the Alexopoulos study, venlafaxine in Montgomery 2004). The confidence intervals for the odds ratios for all the studies pass through 1, indicating non-significance for the efficacy for escitalopram. However the pooled odds ratio has tighter confidence limits that do not pass unity and therefore suggests that escitalopram has an improved effect over its comparators. Whether

an OR of 1.32 can be regarded as an indication of a strong effect is perhaps questionable when trying to apply this to clinical situations. Tests for heterogeneity indicated there are none (Cochran Q = 3.7, df = 5, P = 0.59).

Discussion

There is little doubt that escitalopram has an interesting pharmacology that appears to confer an advantage over its racemate, citalopram. The concept of an allosteric mechanism at the transporter causing a greater than expected response is one that should be explored further to perhaps produce more effective antidepressants.

This study has found that the pre-clinical animal data supports the differential activity of citalopram and escitalopram at the SERT, but it is more difficult to translate this effect into human terms; that is, a clinical effect. It does seem that patients need only half the dose of escitalopram as citalopram; i.e. 10mg escitalopram is equivalent to 20mg citalopram. This results from the activity residing in the S-enantiomer with little or none in the R-enantiomer, but the latter, through the allosteric receptor site on the transporter, prevents full expression of the efficacy of the S-enantiomer.

The efficacy and side-effect profile of CIT are well established (Bouchard *et al.* 1987; Ekselius *et al.* 1997; Patris *et al.* 1996; Rosenberg *et al.* 1994; Shaw *et al.* 1986; Stahl *et al.* 1998; Elsborg 1991) and, since ESC is one-half of the racemic mixture CIT and shown in animal experiments to be the active enantiomer, it should follow that ESC should be at least as efficacious as CIT and with at least the same good side-effect profile. At the time ESC was launched, the data provided for clinicians was relatively sparse: only five clinical reports were available, mostly as posters from conferences, plus a pooled analysis of three trials (Gorman 2002), one of which was not available in the public domain at that time (subsequently published as Burke *et al.* 2002)

As citalopram is an effective and well-tolerated antidepressant, this begs the question as to why a prescriber would want to change a patient already receiving citalopram to escitalopram, although changing from one SSRI to another can be beneficial. This is borne out in the open-label Zimbroff *et al.* study (2004) in which patients were switched from an SSRI to escitalopram. There appears to be little difference in tolerability between CIT and ESC, although there were twice as many dropouts in the escitalopram group in the failed MD-02 trial. Efficacy is similar, although the trials to date all favour ESC over CIT, except MD-02. The meta-analysis presented here appears to support this.

However, in the individual studies, there is little statistical difference between escitalopram and the comparators.

Across all the studies, there was no difference in tolerability, with almost half the adverse events occurring in the first week. Headache was the most common event in each group. The time to response was reported in only five of the studies, giving little data to perform a comprehensive analysis (Table 22). The times to response were estimated from the graphs representing the change from baseline to endpoint for MADRS score in the study reports. Lepola *et al.* (2003) did report the time to response in their study based on median survival times: 8.1 days faster for escitalopram-treated patients than for citalopram-treated patients. The estimate made from the graph in their paper agrees with their estimate (Table 22). Generally, escitalopram has a faster onset of action except in the study by Bielski *et al.* (2004) in which venlafaxine appears to have the more rapid time to response. Overall there is an advantage in favour of escitalopram of 7-10 days. In terms of drop-outs, the data again is not very comprehensive but there may be a trend towards escitalopram having a better drop-out rate, although the data is equivocal. Using the higher dose of escitalopram (20mg) does seem to lead to more adverse events than the lower dose (10mg), so that there is no difference between escitalopram 20mg and citalopram 40mg (Burke *et al.* 2002).

When the pooled analysis by Gorman, which was a key paper quoted in the early launch literature was analysed in detail for this study, it does not convincingly argue in favour of escitalopram. A similar conclusion was reached in Clinical Guideline CG23 published by NICE in November 2004. NICE has analysed only the Burke, Montgomery, and Wade trials for placebo-controlled data. For comparison against other antidepressants, they used Bielski for comparison against venlafaxine (Bielski *et al.* 2004), for CIT as a comparator the Montgomery data was used (2001), and Alexopoulos (2004) who used sertraline as the active comparator.

In the NICE analysis, there is some evidence that ESC is statistically better than placebo for reducing depression symptoms as measured by the MADRS but the size of the difference is unlikely to be of clinical significance. Against SSRIs or venlafaxine, NICE state that there is either insufficient evidence to determine whether a clinically significant difference exists or that there is a suggestion of statistical difference but not clinical difference from the available evidence. As can be seen in Table 21, the MADRS score change from baseline to endpoint is quite small.

One study in the FDA submission is described as a 'failed trial'. MD-02 does not support Lundbeck's claim that escitalopram is indicated for major depressive disorder but it does not reject it either. This trial was conducted in similar manner to another trial submitted to the FDA (MD-01, later published as Burke *et al.* 2002), although the age range was higher, allowing ages up to 85 years: MD-01 (as with most trials) allowed ages only up to 65 years. The FDA report compares the mean ages in the two studies: MD-01 had a mean age of 40±12 years with 6% of subjects ≥ 60 years old, while MD-02 had a mean age of 42 ±12 years with 9% of subjects ≥ 60 years old. There was also an apparently larger placebo effect in MD-02 compared to MD-01. All other studies, both for the submission and subsequently, have not had similar findings. Without further detail than that given in the submission summary, it is difficult to determine why there is this equivocal result for this trial, as the age range is very similar to that found in other randomised controlled trials, although the percentage of subjects ≥ 60 years old is slightly higher than in other trials.

This thesis has added to the four meta-analyses that were available up to early 2006. The reason for this re-analysis was to attempt to provide an unbiased analysis. Out of these five pooled and systematic studies, only one was not supported in any way by Lundbeck: this was Svensson and Mansfield (2004), which was negative in its conclusions. The Svensson and Mansfield study (2004) is a systematic review and lends no extra analytic view on the efficacy of escitalopram.

As this thesis was being finalised, another meta-analysis was published (Kennedy *et al.* 2006). The authors performed their analysis using 'original data from patients who participated in all MDD studies... that directly compared escitalopram with other antidepressants'. They state in their method that raw data from each patient were used entered into the analysis. One of these studies, however, is a trial that compared escitalopram with paroxetine in the treatment of generalised anxiety, while another presented data on the use of escitalopram in the elderly depressed. They performed meta-analyses on the overall population and one on the severely depressed population, with analysis for response to treatment and remission rate, expressed as odds ratios. Escitalopram was found to have greater efficacy in the overall population, as assessed by the decrease in MADRS score from baseline. However, the improvement was only 1.22 MADRS points greater than with conventional SSRIs, a difference which is not regarded as clinically significant, even if it is statistically significant. The overall odds ratio for response was estimated to be 1.29 (95% CI 1.07-1.56, P=0.35); for remission it was 1.21

(95% CI 1.01-1.46, P=0.05). These values are not too dissimilar from those presented in this thesis. This author contends that these results show that escitalopram has little advantage over comparators, although Kennedy *et al.* conclude differently. It is perhaps of note that one author was an employee of Lundbeck at the time.

Although analysis reveals escitalopram to have greater efficacy than active comparators and placebo, the differences do appear to be marginal: certainly not enough to warrant the expenditure by the NHS when citalopram is now available as a cheap generic. So why is there this need to prove escitalopram is better than citalopram? It is well known that pharmaceutical companies will produce product extensions or chiral versions of drug entities that are nearing their patent expiry (Angell 2005). It is possible that Lundbeck has done this. There is also the recurring theme with all these studies: apart from the Svensson study, they all have either support from Lundbeck in some form and/or one or more authors were Lundbeck employees.

CHAPTER 6

GENERAL DISCUSSION

Depression has a significant burden of morbidity and mortality with important consequences for the individual and society. The World Health Organisation has projected that depression will be the second highest morbidity globally by 2012 (Murray 1997). The Clinical Guideline published by NICE in 2004 for depression, and the update in 2009, was a major step forward for promoting best practice based on available evidence but unfortunately the resources did not always appear to follow the guidelines.

Evidence-based practice relies on robust data which is gained from randomised controlled trials and meta-analyses. Randomised controlled trials enable the introduction of a new drug onto the market, but they do not in themselves help in deciding if that drug will be a cost-effective option. The more difficult question is how effective drugs are in the real world, as RCTs do not reflect the real world clinical situation.

With the ever-increasing costs in providing health services and an ever-increasing demand for those services, some form of rationing is necessary. Whatever the healthcare system might be, the problem is still the same: how to use scarce resources effectively. New treatments are often considered to be improvements on previous therapies. However this is debatable (Angell 2005) and they often cost more, although pricing of new antidepressants can be competitive.

However, in the case of the antidepressants, this rise in costs with new drugs has not generally been associated with improvement in treatment efficacy or effectiveness to justify them. Drug development is still rooted in the past, developing or re-engineering drug molecules based on old mechanisms of action, which gives rise to drugs that have no greater efficacy than earlier ones, albeit with perhaps cleaner, safer adverse event profiles, but which are essentially 'me-too' drugs. In the early 1990s in Australia, antidepressant costs were escalating in spite of prescription numbers remaining constant (Alchin and Tranby 1994). TCAs were still first-line choice, but second generation drugs -selective serotonin reuptake inhibitors (SSRIs), tetracyclics and reversible inhibitors of monoamine oxidase A (RIMAs) were gaining popularity, which had significant acquisition cost implications. For example, in Australia, at 1994 prices, amitriptyline only cost A$5.84 per prescription, compared with A$45.39 for moclobemide (Alchin and Tranby 1994). The prescription length was not specified in Alchin and Tranby's paper, and it is also not clear what doses are being considered; the TCA may or may not be a therapeutic dose. General practice doctors have been known to prescribe what are generally considered by hospital psychiatrists to be subtherapeutic doses (Blacker and Clare 1987; Johnson 1974; Brugha *et al.* 1992; Donoghue & Taylor 2000). Although TCA costs are usually low, if they are being prescribed in subtherapeutic doses this is a waste of resources, and the ensuing

economic burden is large, irrespective of the acquisition cost of the drug. In addition, there may be some morbidity due to side effects without any treatment benefit. Several of the SSRIs can be prescribed in therapeutic doses from the start of treatment. For example, the recommended treatment and maintenance dose for fluoxetine is 20mg ('Prozac' Sept 2005). Dose increases for the newer drugs are not generally necessary as the starting dose is usually considered the maintenance dose, particularly in primary care, although in hospital settings this may not be the case and dose increases in moderate to severe depressive illness are not uncommon. Mirtazapine, sertraline and venlafaxine usually need dose increases. Although the lack of dose titration in primary care makes prescribing easier and less complicated for the GP, it can escalate costs when using newer technologies until generic versions are available. However, pharmacoeconomic analyses can demonstrate a significant cost offset when the costs associated with absenteeism or presenteeism, for example, are taken into account. The situation is further complicated by the differences in pricing between the primary and secondary sectors, with hospitals usually enjoying a greater discount on purchasing than the community, although in primary care VAT is not charged.

In 1994, SSRI prescribing was increasing and accounted for approximately 15% of antidepressant treatment costs while contributing to 50% of total costs of NHS antidepressant prescribing (Gilchrist and Knapp 1994). By mid-2001, 5.9 million items for antidepressants were dispensed at a cost of £82.3 million, of which SSRIs accounted for 2.9 million items at a cost of £55.5 million. From 1996, SSRI prescribing increased by 143% while the cost rose by 66%. The influence of branded products going generic during this time is exemplified by fluoxetine, which is the most frequently prescribed SSRI. Since fluoxetine came off patent the cost per pack has fallen from the initial launch price, so that the drug cost fell to £11.3 million (for 988,000 items) per quarter, due to the prescription and dispensing of cheaper generic fluoxetine. Compare this to paroxetine over the same period, which had fewer prescriptions (933,000 items) but with costs rising to £22.9 million per quarter (PPA 2001).

Pathophysiology & Pharmacology

This study has highlighted the discussion about the disease itself, particularly whether the traditional theory of dysregulation of the monoamines are the cause, or whether there are other factors that contribute more to the development of the illness. It is more likely to be a combination of factors or events that lead to depressive illness. The influence of high steroid levels in the brain possibly as a result of stress, dysregulation of

intracellular mechanisms of repair, and disturbance of transcription factors all seem to play a role in the dysfunction of monoamine neurotransmitter systems, presumable post-synaptically, leading to what we see clinically as depression. Until a better understanding is reached regarding the genesis of the illness, it will be difficult to develop better therapies, particularly biological.

The action of antidepressants is under scrutiny as well, as there is a difference of opinion concerning the role of the so-called 'dual action' agents: venlafaxine, milnacipran, duloxetine and mirtazapine. There is ongoing debate as to whether dual action is necessary for more rapid onset of action and greater efficacy, or whether drugs with a more specific mode of action are just as effective. It has been suggested that the SSRIs, currently the mainstay of treatment of depression, may in the future have only a role in treating anxiety and some subtypes of depression (Taylor & Stein 2006). Evidence suggests they are only moderately more effective than placebo in treating major depressive disorder, and that they are probably only useful for certain types of depression. The different effects of the various antidepressants seen in clinical practice between different patients (i.e. one might not work but another will) may be due to these differing forms of depressive illness, although one study that compared trials of reboxetine and fluoxetine found that there was no reason to believe that symptom differences (which may indicate differing depressions) are useful for antidepressant selection (Nelson *et al.* 2005). The hypothesis is tentatively supported only if depression is defined according to genetic factors, and not when the definitions are either for more rather than less severity of illness or for melancholic versus non-melancholic depression (Taylor & Stein 2006). However, it may explain the belief that some psychiatrists have that SSRIs are not always that effective. Equally, this could well derive from the type of patients seen by mental health services. They will normally have been seen by general practitioners and referred on because of illness that has become difficult to treat in primary care.

Rating Scales

Measurement of depression is by using rating scales. These constructs of the symptoms and severity of the illness are open to some degree of interpretation and for clinical trials raters have to be trained how to use them. This brings into question the degree of accuracy between raters, and the trial reports reviewed for this study do not always state if raters have been specifically trained in the correct use of the chosen scales. The main problem of using rating scales is that they are surrogate measures: they do not objectively and unequivocally measure the symptoms, unlike 'hard' outcomes such as

blood pressure values or biochemical parameters. Measuring the depth of depression is a 'soft' science: no hard data can be elucidated. So it is inevitable that there may be some subjectivity, no matter how objective one tries to be.

This thesis has reviewed the two principal rating scales, the Hamilton and Montgomery-Åsberg, which are used in randomised controlled trials. This review gives a sense that they are probably outdated and should be replaced by a scale that reflects modern living. As the two diagnostic classification systems. ICD-10 and DSM-IV, are being revised and harmonised, it may be opportune to revisit the rating scales. This author contends that there should be a closer match between the scales and the classification systems but that any new scales must focus more on the main diagnostic features and symptoms of depression. A debate is needed as to what actually constitutes the prime symptoms of depression and what are secondary symptoms that might equally be attributable to other illnesses, environment or drug therapy. Sleep disturbance, for example, may be attributable to environmental reasons, other drug therapy or a stressful situation in which the sufferer is unable to cope easily and not therefore as a result of a depressive illness. Other symptoms such as appetite changes, alterations in libido, even anxiety, can be attributed to non-depressive illness or stressful circumstances. The development of the HADS and EPNDS were designed with this concern in mind. It is this type of diagnostic differentiation that needs consideration. Such a debate is outside the scope if this thesis but is important future work (David Taylor, personal communication, 2010)

Clinical Trials

Clinical trials are of necessity short-term (usually up to six weeks) to prove efficacy for licence submission and therefore the primary outcome required is most likely to be response. However remission should be considered more important as this is the long-term goal clinically, which shows the drug will reduce symptoms to zero or to a minimum. The true outcome in treating an illness is removal of disease. (That is cure, or complete remission that is at least long lasting but preferably permanent: remission being the state of having no symptoms although the disease is still present.) An example of a fully 'remitted' disease state (cure) would be that resulting from an appendectomy: the patient would never have appendicitis again. However in mental illness it is rarer for a patient to be entirely free of illness after initial remission. These can be enduring illnesses, although some patients may only have one episode in their lifetime. So the best achievement is a surrogate outcome: reduction of the disease to a manageable level by the use of interventions that use

intermediate outcomes as a measure of efficacy. In depression, particularly in clinical trials, severity is estimated at different time points. In RCTs, the achievement of remission is not usually defined as complete resolution of symptoms and subjects may have residual symptoms. It is probable that the time span of trials may be too short to elucidate this. In terms of translating the resolution of symptoms in a randomised controlled trial to the ability of a patient able to function socially and in the workplace, the standard trial requirement of a 50% reduction in total rating scale score for response does not necessarily indicate this ability. It only shows that there is a certain degree of improvement at a particular time point (Fawcett & Barkin 1997).

A further issue with these estimates of efficacy is the difference between the investigational drug (I) and either the standard drug (S) or placebo (P). Ideally, I>S or P but it is usually the case that while I>P, I=S. Unfortunately, due to placebo responses becoming more pronounced, this difference (the effect size) is getting less (Schatzberg & Kraemer 2000). The difference in raw scores between I and P on either the Hamilton Depression rating scale (HAMD) (Hamilton 1960; 1967) or Montgomery-Åsberg Depression rating Scale (MADRS) (Montgomery & Åsberg 1979) can be relatively small; there should be at least a 3-point difference to be clinically relevant. When calculated, effect sizes should be large ($\geq$0.8) to indicate that I has a statistically significant effect (Keck *et al.* 2000). Small effect sizes, which indicate the active treatment has little or no advantage over placebo or active comparator, are becoming more common, and can result in 'failed' trials.

Drug response rates have remained relatively stable, even reduced. This may be a result of using more resistant populations. There is also a high drop out rate from studies, skewing results, particularly if the drop out occurs early in the study. Last observation carried forward (LOCF) is a commonly used method of extrapolating data to fill in later missing data points. Low early scores may thus be carried forward and reduce the end result. Protagonists argue that this is a fair test and militates against a falsely raised positive result. However, a patient who has dropped out due to side effects may have had higher rating scores at later time points, if they could have continued, possibly even finishing the trial. Patients with better tolerance to side effects are more likely to do this. Such patients may have a different sub-type of depression to others, as they may have a severe depression that motivates them to stay in the trial, in spite of unpleasant adverse effects. Conversely, those patients with mild depression will possibly tolerate even mild side effects less and drop out. This makes the generalisation of the results of trials to the general population difficult, particularly if the drop out rate is greater than 30% of the study

population (Montgomery 1999b). Conversely, if a true end-point analysis was to be used for all patients who completed, there is the possibility that enough low end-point scores would cause the trial to 'fail'. Although completer analysis is used, it is rarely clear that all scores have been utilised and there is the possibility that the results from patients who withdraw early may be excluded (Moncrieff & Kirsch 2005).

Randomised controlled trials can only decide the efficacy of a drug but in clinical practice it is effectiveness that is important. This has seen an increase in the conduct of real world clinical trials. Intervention intensity is potentially less than in RCTs as patient care should follow usual practise as closely as possible, while the control group may be 'care as usual', which often involves active treatments. The severity of the depressive episode may not be evened out in a real world trial as it might in a typical randomised controlled trial. In both types of trial, rigorous assessment of the depth of depression is essential.

Randomised Controlled Trials

The standard method of evaluating antidepressant drugs is by conducting a randomised, double blind, placebo-controlled trial, particularly for new antidepressant agents. For registration purposes, one placebo-controlled trial can provide enough evidence to prove efficacy, although it is more usual to have two or three (Montgomery 1999a). More than this number and it becomes ethically challengeable: such methods raise ethical issues regarding the use of placebos, blinding, the patient population included, industry sponsorship, the way in which trials are reported, the rating scales used. There are also methodological problems associated with the measurement of the depth of depression, the response to placebo, and how to determine the points at which response and then remission are reached. There are no precise measurements, only rating scales which are potentially subjective. Randomised controlled trials can include the investigational compound being compared against active controls, with or without placebo. But these data need to be translated into effectiveness in the real world, which requires either large prospective trials (difficult and expensive to conduct) or scenario modelling using data derived from clinical trials and other sources.

Patients and investigators should be blinded to the treatment being received, there should be defined criteria for entry into the trial, validated rating scales and/or other instruments should be used for evaluating the severity of depression and the course of improvement, outcome criteria should be defined at the start of the trial, and therefore ambiguities and biases should be virtually eliminated.

However, reality dictates otherwise. Blinding is not always perfect, as the lack of response is often interpreted by both parties as receipt of placebo when, in fact, it could be due to a true lack of response to active compound. This does not, however, infer the drug has no activity overall, only that an individual may have a physiological disposition such that that particular molecular entity would have had no activity anyway. The drug may well have efficacy in other subjects who have a different physiology in brain connectivity. There may also be side effects from the active compound, which identify the subject as receiving it.

The Placebo Effect

To compound these problems, there is evidence to suggest that the placebo arm cannot be considered as inert and inactive. Large placebo responses have been noted in some trials, for example, averaging 29% in studies of acute bipolar disorder (Keck *et al.* 2000). This effect results from using an intervention that produces a therapeutic effect by intent and not as a result of its pharmacology. This poses problems for clinical trials as it becomes difficult to dissociate true response to active drug from the response due to placebo. From the regulators' viewpoint, randomised, double blind, placebo-controlled trials have become the gold standard by which new drugs are evaluated. But meta-analyses have shown that placebos can duplicate the active drug response by 65-80% (Kirsch 2000).

This response in the placebo arm may be due to a number of factors. First, the subjects in the placebo arm are not receiving a null intervention. They are being intensively monitored, possibly by more than one person, and therefore receive non-pharmacological intervention in the form of support and encouragement during the course of a clinical trial (Schatzberg & Kraemer 2000). There is an expectation of being treated. Second, following on from this argument, some patients might have a depression that is more responsive to psychological intervention. Third, personality may define subjects as being responsive to a placebo, since they consider themselves fortunate to be receiving therapy, albeit placebo, they would not otherwise have had if they had not participated in the trial (Mattocks & Horowitz 2000). Personality and temperament may therefore influence the disposition of a trial subject: a desire to be included, more likely to respond to placebo, and possibly different responses to rating scales.

However, the mistake must not be made in asserting that the placebo arm is as good as the active treatment, thereby rendering placebo as not necessary (Lavori 2000). Some argue that it is unethical to use placebo and suggest that either active comparators or active placebos should be used. The use of placebos raises the issues of consent and also of

giving a non-pharmacological treatment to patients who may suffer harm as a result of not receiving active treatment.

For a placebo to be effective in disguising its true nature, blinding must ensure that neither the investigator nor the subject can differentiate placebo from active. There may be various clues, for example lack of side effects, lack of efficacy and, although guesses may be inaccurate, the majority of doctors and patients will guess correctly (Rabkin *et al.* 1986). Active placebos, which mimic some or all of the side-effect profile of the active investigational drug without any inherent psychotropic activity, have been suggested as a method of overcoming these problems. However, placebos also have side-effects that can mimic those of the active compound, and the effects (pseudotherapeutic and adverse) of a placebo need to be elucidated before the effects of active treatments in RCTs can be assessed (Weirauch and Gauler 1999).

The major issue is placebo response, rates of which appear to be rising (Montgomery 1999b). This response is usually taken to mean the apparent improvement in the clinical state of patients who have been assigned to the placebo arm. It does not describe the efficacy of placebo but does show that a subgroup of the patient population has an improvement in their symptoms. This may relate to the attitude and temperament of the patients. The relapse rate is high, however, when compared to the true improvement gained from active treatment.

Severity of the illness is one factor that may be influencing this large response from placebo. It appears that patients who are less severely depressed respond to placebo more readily, whereas those with a longer duration of illness have fewer tendencies to show a placebo response (Fairchild *et al.* 1986). However, too long a duration of illness could mean the patient is treatment-resistant, which would not be a fair test for a new compound. The optimum duration for illness to avoid these issues has yet to be clarified.

In this study, no evidence was found to suggest that authors of RCTs allowed for the placebo effect. Placebo controls are currently a necessary evil. It is difficult to see a viable, ethical alternative that will satisfy regulatory authorities and ethicists. Equivalence studies can demonstrate that the investigational drug has the same degree of efficacy as the standard drug but these active comparator studies have no internal validation without a placebo arm. This can lead to 'failed' studies such as the Lundbeck MD-02 study of escitalopram vs citalopram and placebo in which there was no significant difference between them on the primary outcome measure.

Data Analysis in RCTs

In an ideal world, all subjects would complete a trial: i.e. there would be no dropouts due to any reason. However, patients may withdraw from trials due to lack of efficacy, intolerance of side effects, loss to follow-up for other reasons, or violation of trial protocol. Therefore only a proportion of subjects will have a complete set of data points (i.e. they attended all scheduled visits). Another group of subjects will have dropped out early (perhaps after only one post-baseline visit). These patients will have no further information to contribute to the analysis after their last visit, so the data collected at that point will be used for endpoint analysis: i.e. last observation carried forward (LOCF). The use of LOCF introduces biases, however, as the early data do not necessarily predict the outcome at trial endpoint. We cannot say that patients' responses would have remained constant after the last data set and that the values in that data set would be those at endpoint if the subjects had completed the trial. This therefore lowers the estimate of effect, giving a more conservative estimate of the efficacy, which some researchers say argues in favour of the technique. Conversely in some situations, LOCF may actually overestimate the treatment effect, usually when there is a high dropout rate in the comparator group, while it may underestimate the inferiority of the inferior treatment (Mallinckrodt *et al.* 2003). LOCF may also exaggerate the size of effect and increase Type I error (i.e. falsely conclude a difference exists when in fact the difference is zero).

A method by which these missing data are handled more precisely is needed, although some might argue that LOCF has been used for many years and is well known. There are three situations in which data can be missing. *Missing completely at random* (MCAR) arises when the missingness is not explained by the outcomes of interest being either observed or unobserved. If the missingness is explained by the observed outcomes but not the unobserved outcomes, the data is said to be *missing at random* (MAR). If the converse of the latter is true, that is missingness depends on the unobserved outcomes, the data is *missing not at random* (MNAR). LOCF data are considered to be MCAR: subjects' responses would remain constant from the last data point to the endpoint. Such observations may therefore bias the estimate of treatment effects and the associated standard errors.

A further complication of LOCF is that it is effectively a snapshot of drug performance. Although there is obviously a time period over which observations are conducted, there is a clear objective in that the endpoint is the measurable effect. However, the intervening observations are not considered, so the profile of treatment effect during the time of observations is not considered or accounted for.

Eli Lilly and Company has recently used repeated measures analysis to overcome these shortcomings in LOCF for the new SNRI, duloxetine (Detke *et al.* 2002a; 2002b; Goldstein *et al.* 2002; 2004; Brannan *et al.* 2005; Burt *et al.* 2005). Mallinckrodt and co-workers have developed the mixed-effects model repeated-measures analysis (MMRM), which estimates the treatment effect over time, accounting for early dropouts (Mallinckrodt *et al.* 2001; 2003), for the particular requirements of acute-phase clinical trials. Scrutiny of the duloxetine RCTs revealed references to repeated measures analysis, which has been around for some time: a further search on PubMed revealed 78 references using the search term likelihood-based mixed-effects model, with the earliest being 1984.

It remains to be seen if other researchers and pharmaceutical companies take up this method of analysing trial data but it does appear to be a method that might avoid some of the problems associated with LOCF and observed case analyses. It is noted in this study that Colonna *et al.* (2005) have used a form of repeated measure analysis of variance as their primary measure of antidepressant efficacy, while all other investigators have used the standard method of LOCF. The author of this current study believes that MMRM should be investigated further for its potential in psychiatric drug research.

Meta-analysis

This quantitative method of pooling studies, i.e. an analysis of data already analysed, is not the same as reanalysing the primary data from the individual studies. It can be used to confirm the findings from the original studies or to answer new questions arising from those studies (Noble 2006). New questions might include the effects within subgroups (eg male vs female) or the incidence of side effects. Obtaining the original data from all trials would be the preferred method but meta-analysis is usually the only option available as obtaining the original data from the authors is difficult. It would be interesting, for example, to re-analyse the escitalopram raw data using MMRM. However few, if any, published studies include raw data or enough data to enable calculation of odds ratios for subgroups or for side effects (Thompson & Higgins 2005). The ideal would be to conduct trials with large numbers of subjects: these would be less subject to chance findings and would possibly reduce the need for meta-analysis. Small studies are potentially more prone to chance; hence the need to combine them and create a larger patient population to increase the statistical power, as has been done recently in a paper analysing data from many studies of different antidepressants and cross-comparing them against each other (Cipriani *et al.* 2009).

As in original research, meta-analysis requires a specific question to be addressed that will be answered by searching for appropriate original studies (usually RCTs), analysing the studies for methods and population, and analysing the data so derived. In the current analysis of escitalopram described in this thesis, all the available randomised controlled trials have been included, discarding the open label and extension trials: different trial formats should not be mixed in meta-analyses. Care was taken to ensure that the subject populations were as matched as possible in terms of age, inclusion and exclusion criteria, and severity of depression at baseline. Large differences between trial populations could introduce bias. The primary outcome measure was the same in all studies, using the MADRS as standard; if the Hamilton Depression Rating Scale had been used in some trials as the primary measure, this could have introduced further bias as the constructs of the two rating scales are based on slightly different conceptual ideas of depression.

On balance, the result from the meta-analysis of the efficacy data from these studies shows that there is a small effect in favour of escitalopram but this author's contention is that the effect does not warrant inclusion in a formulary when the cost is taken into account. However, this only applies when considering the cheaper alternatives, such as citalopram or fluoxetine. If venlafaxine and its associated cost for most brands are factored in (venlafaxine is a commonly-used antidepressant), then the argument does seem to favour escitalopram. For PCTs and acute trust formularies, the argument for exclusion centres on citalopram having as good efficacy and tolerability but far lower cost.

To overcome the problems associated with meta-analyses, the ideal would be to conduct original RCTs that have large numbers of subjects. However, such trials are expensive and are logistically very complicated, and the sample sizes involved could pose ethical problems, particularly if placebo is involved (Noble 2006).

Pharmacoeconomics

The burden of depression is large, set to become the second major morbidity by 2012. Although acquisition costs for antidepressants are relatively low when compared with some other treatments (eg cancer chemotherapy), the total cost is high due to the increasing prevalence of depressive illness.

This study has reviewed two aspects of the economy of the illness: the cost of depression and the cost of treating it with drug therapy. Escitalopram has been used as an example of a recently-introduced antidepressant drug. Drugs and therapeutics committees (medicines management committees) in the UK have not universally accepted it for

formulary inclusion as the evidence from RCTs is not convincing enough to justify the cost: a view also borne out by this study. This author is of the opinion that this can be a narrow view and that the cost of not treating with a new compound may actually have cost consequences for the health economy by not taking into account the non-pharmaceutical costs.

The current review reveals that between 1993 and 2003, the cost to the UK of treating depression rose from over £3.3m to a little over £9m, during which time the drug costs rose approximately six-fold. This increase in cost may well be due to the introduction of new antidepressants with higher acquisition costs than the older, established, drugs.

Where escitalopram is concerned, several cost-effectiveness analyses were published, all suggesting that the drug is a good cost-effective choice for clinical practice. Modelling studies also bear this out. So the overall conclusion appears to be that escitalopram is a cost-effective alternative to other antidepressants, even taking into account the higher purchase cost. However, there needs to be some caution as many of the various studies are industry sponsored or authored.

Industry Influence

Unfortunately, there are issues with the data on which the various pharmacoeconomic and clinical research studies have been based. Randomised controlled trials for licence submission are, by necessity, run by the pharmaceutical industry. Therefore the literature on antidepressants is under the control of the pharmaceutical companies and is potentially flawed. Trials produce multiple data that can be manipulated to produce desired outcomes, but these secondary data sets are not derived from a priori hypotheses. By this 'data dredging', researchers can produce results to prove whatever is wanted (Procopio 2005), while the multiple statistical analyses may result in Type I error (i.e. a false positive result) (Taylor & Stein 2006).

Additionally, and as stated previously, the differences between various antidepressants and/or placebo can be small, although trials of new drugs invariably show that the new drug is better than the comparators by a (usually) small margin. Given the potential for manipulating the results, it could be that the drugs are very similar to each other. It does not help that drug trials are based on non-falsifiable hypotheses. (Hypotheses can be 'falsifiable' or 'non-falsifiable'. In the former, the hypothesis can be proven false but it can never be demonstrated to be true. Non-falsifiable hypotheses cannot even be proven false and therefore will always have a degree of uncertainty about them (Procopio 2005))

The funding of studies invariably leads to clinicians and healthcare providers being sceptical about the results. If healthcare providers, including government bodies, want a clear picture of a new drug's position in treatment, they will need to fund independent studies. In the UK, NICE is such an organisation (although it does not fund studies), allegedly independent of Government, which appraises new technologies including drugs, but this author contends that even NICE might find it difficult to separate itself from Government influence. To have true independence of any interested parties is probably virtually impossible, as the organisation providing the funding will want a result it wishes to have. The Cochrane Collaboration also reviews new health technologies and may be more independent of governmental and industrial influence, although even their meta-analyses are potentially contaminated, being based on industry-produced trial reports (Noble 2006).

The pharmaceutical industry is global, composed of a few large multinational companies, and a small number of independent ones, which includes the biotech industry. In many fields of medicine, not just depression, the industry has relied on the innovative research of university and other independent research organisations to produce the advances in drug treatment (Angell 2005). With the possibility that the literature is at least in part controlled by the industry, it is important to ensure that the reporting of RCTs is accurate and somehow uninfluenced by industry. Accuracy has been promoted by the use of the CONSORT statement (Moher *et al.* 2001), although not all authors follow all recommendations. This study has used the CONSORT criteria to assess the published RCTs for escitalopram, and only Moore *et al.* (2005) have closely followed these guidelines. However, many trial reports do not predefine outcome measures or statistical power very often. The declaration of sponsorship appears to be much better reported.

Conclusions

Depression is a common and complex illness to treat that is still treated with drugs, although there is a move in the UK to treat mild depression with psychological therapies. The constructs of depressive illness have changed over the decades and there has been a realisation that personality, genetics, and environmental factors all play a role in determining the onset of depressive illness, its continuing course, and its resolution.

This thesis has attempted to draw together several elements of the process involved in the development, treatment, and economic aspects of the pharmacotherapy of depression. The early monoamine theory of depression has been used as the basis for developing drug treatments since the 1950s and, in spite of experimental work to

investigate the underlying mechanisms of depressive illness (and thereby develop new drug interventions), there have been no novel chemical entities developed and brought to market other than those that increase the monoamines noradrenaline and serotonin. It can thus be said that the available drugs are all essentially 'me-too' compounds.

However, drug treatment is the mainstay of treating mental illness. Depression has a higher prevalence of some 10-12% than the other major mental illnesses (schizophrenia, 1%; bipolar disorder, 5%). Therefore, although the acquisition costs of antidepressants tend to be low in comparison to those of the atypical antipsychotics, because of the number of potential patients involved, the cost of treatment with antidepressant drugs will be substantial. This makes the economics of treatment important to consider. On balance, it does appear that antidepressant treatment can offset the more intangible costs related to the illness, such as loss of earnings, costs to society. Unfortunately health care providers, particularly the National Health Service in the UK, do not appear to accept the findings of researchers in this field, and continue to insist that it is the drugs budget that needs to be reduced by using lower cost alternatives (i.e. cost minimisation). Although this will relieve the financial pressure on the NHS in the short-term, based on the current understanding of pharmacoeconomics, it is a false premise.

Because of the reliance on antidepressants to treat depressive illness, it is essential that the drugs available are effective and safe. RCTs are designed to answer these issues, although long-term safety will not be addressed in the usually short duration of RCTs. Treatment effect is evaluated by measuring the differences between the investigational drug and active comparators or placebo in scores on depression rating scales. The difference should be statistically significant and both individual RCTs and meta-analyses have demonstrated this. However, the odds ratios derived from such analyses maybe quite marginal: this study shows that for escitalopram, the odds ratio is only 1.32 (95%CI = 1.07-1.63) for remitters in the active comparator group. Although statistically significant (P = 0.013), this odds ratio suggests escitalopram is only reasonably more likely to produce a statistically greater effect than a comparator. Clinical effect is perhaps less significant.

Although RCTs are essential to evaluate new drugs (in spite of their apparent drawbacks), the real world randomised trial may be the way to prove effectiveness in clinical practice. RCTs have 'pure' cases (i.e. patients who are carefully screened to eliminate potential confounding co-morbidities), unlike real world trials with 'ordinary patients', who may have other mental or physical illnesses. Due to the greater variation in the 'wild', it may be more difficult to detect a meaningful difference between treatment groups; such a difference is usually taken to mean a statistical power of 0.8 (Keck *et al.*

2000). In the real world, the greater variation found in the population of the treatment arms of a trial will reduce the statistical power and the precision of a statistical estimate for a given sample size. Therefore real world studies will need larger sample sizes to achieve the same power and statistical validity: a priori sample size and power calculations should be performed (Sturm *et al.* 1999) to determine this, and selection of the population sample needs careful control.

Parametric sample selection has been the most widely used method in observational and retrospective depression studies, although these have been few. Using these techniques, studies have shown that the higher acquisition costs of newer antidepressants are offset or more than offset, by 'broader measures of healthcare utilisation' (Crown 2001). These techniques reduce biases inherent in non-randomised trials, or in quasi-randomised controlled trials.

The combination of data from RCTs and real world trials may reinforce findings as the observational data if consistent, help to inform the understanding of the economic outcomes of alternative treatment strategies. The process of combining data from RCT and economic studies should be pursued to unify the real world development of drug therapy after pre-clinical studies have shown the new chemical entity to be efficacious. This would help inform our understanding of its effectiveness. A prospective economic analysis run parallel to trials (either RCT or real world) would complete the assessment of a new drug's position clinically and in the market.

Current rating scales for depression used in RCTs (i.e. the Hamilton and the Montgomery-Åsberg) may now be less sophisticated for the information requirements for today's clinical needs. This study has indicated that, of these two particular scales, the MADRS is possibly the more useful as it is less influenced by somatic symptoms that may result from side effects of drugs. In particular, the Hamilton 17- or 21-item depression rating scale total score as generally used in RCTs is not based on a unidimensional scale, which reduces its functionality to measure the severity of depression, uninfluenced by symptoms not necessarily associated with depressive illness. A reduced version of the Hamilton scale based on six items, the Hamilton Depression Subscale (HAM-D6), does fulfil criteria for unidimensionality (Licht *et al.* 2005). A rating scale that more closely reflects the root symptoms of the illness and is less influenced by the side effects of drugs should be developed for use in clinical trials that are being conducted for product registration. A social functioning scale and/or daily functioning scale should also be included in trials (particularly real world) to assess the impact of the depressive illness on the individual, particularly to understand the improvement in functioning as depression

lifts. The intention would be to relate the change in depression rating scales to the change in functioning. This may give a more accurate picture as to the effect of the drug on the patient's depressive episode and ability to function. The development of a comprehensive scale, or set of scales, should lead to more meaningful clinical trial outcomes. Pharmacoeconomic analyses based on such data derived from randomised clinical trials may also be more robust.

Finally, the industry needs to develop drugs based on new mechanisms and provide clinical trial data in a more open manner. The currently available therapeutic interventions are all based on the same mechanism: that the monoamines are depleted in serotonergic and noradrenergic brain pathways and by increasing these neurotransmitters relief of depression will be obtained. It is clear that in spite of successes using this basic strategy, not all patients are successfully treated and therefore new drugs that utilise other mechanisms need to be developed and brought into clinical use. The concept of running licensing studies parallel with economic and longer-term safety and effectiveness studies also needs serious consideration.

REFERENCES

Alchin, T. M. and H. W. Tranby (1994). "Who Bears the Cost of Antidepressants in Australia." Medical Journal of Australia **161**(9): 547-549.

Alexopoulos, G. S., J. Gordon, et al. (2004). "A placebo-controlled trial of escitalopram and sertraline in the treatment of major depressive disorder." Neuropsychopharmacology **29**: S87-S87.

Altman, D. G. (1996). "Better reporting of randomised controlled trials: The CONSORT statement - Authors must provide enough information for readers to know how the trial was performed." British Medical Journal **313**(7057): 570-571.

Altman, D. G., K. F. Schulz, et al. (2001). "The revised CONSORT statement for reporting randomized trials: Explanation and elaboration." Annals of Internal Medicine **134**(8): 663-694.

American Psychiatric Association (1994). Diagnostic and Statistical Manual of Mental Disorders, 4th edn (DSM-IV). Washington DC, American Psychiatric Association.

Anand, S. and K. Hanson (1997). "Disability-adjusted life years: a critical review." Journal of Health Economics **16**(6): 685-702.

Anderson, I. M. and B. M. Tomenson (1994). "The Efficacy of Selective Serotonin Reuptake Inhibitors in Depression - a Metaanalysis of Studies against Tricyclic Antidepressants." Journal of Psychopharmacology **8**(4): 238-249.

Anderson, I. M., D. J. Nutt, et al. (2000). "Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 1993 British Association for Psychopharmacology guidelines." J Psychopharmacol **14**(1): 3-20.

Andersson, T. (2004). "Single-isomer drugs - True therapeutic advances." Clinical Pharmacokinetics **43**(5): 279-285.

Angell, M. (2005). The Truth about the Drug Companies. New York, Random House Trade Paperbacks.

Angst, J. (1992). "Epidemiology of Depression." Psychopharmacology **106**: S71-S74.

Anon (1983). "Astra's zimeldine withdrawn because of neurological side effects." Pharm J: 360.

Anton, S. F. and D. A. Revicki (1995). "The Use of Decision-Analysis in the Pharmacoeconomic Evaluation of an Antidepressant - a Cost-Effectiveness Study of Nefazodone." Psychopharmacology Bulletin **31**(2): 249-258.

Antonuccio, D. O., M. Thomas, et al. (1997). "A Cost-Effectiveness Analysis of Cognitive Behavior Therapy and Fluoxetine (Prozac) in the Treatment of Depression." Behavior Therapy **28**(2): 187-210.

Areias, M. E., R. Kumar, et al. (1996). "Comparative incidence of depression in women and men, during pregnancy and after childbirth. Validation of the Edinburgh Postnatal Depression Scale in Portuguese mothers." British Journal of Psychiatry **169**(1): 30-5.

Armstrong, E. P., G. H. Skrepnek, et al. (2005). "A cost-effectiveness analysis of escitalopram and sertraline in the treatment of major depressive disorder." Value in Health **8**(3): 393-393.

Arnesen, T. and E. Nord (1999). "The value of DALY life: problems with ethics and validity of disability adjusted life years." BMJ **319**(7222): 1423-1425.

Aronson, R., H. J. Offman, et al. (1996). "Triiodothyronine augmentation in the

treatment of refractory depression - A meta-analysis." Archives of General Psychiatry **53**(9): 842-848.
Attenburrow, M. J., P. R. Mitter, et al. (2001). "Low-dose citalopram as a 5-HT neuroendocrine probe." Psychopharmacology **155**(3): 323-326.
Auquier, P., S. Robitail, et al. (2003). "Comparison of escitalopram and citalopram efficacy: A meta- analysis." International Journal of Psychiatry in Clinical Practice **7**(4): 259-268.
Aursnes, I., I. Tvete, et al. (2005). "Suicide attempts in clinical trials with paroxetine randomised against placebo." BMC Medicine **3**(1): 14.
Åsberg, M., Kragh-Sørensen, et al. (1973). "International reliability and communicability of a rating scale for depression." Psychological Medicine **3**: 458-465.
Åsberg, M., S. A. Montgomery, et al. (1978). "A Comprehensive Psychopathological Rating Scale." Acta psychitr scand **Suppl 271**: 5-27.
Baker, C. B., M. T. Johnsrud, et al. (2003). "Quantitative analysis of sponsorship bias in economic studies of antidepressants." British Journal of Psychiatry **183**: 498-506.
Beasley, C. M., B. E. Dornseif, et al. (1991). "Fluoxetine and Suicide - a Metaanalysis of Controlled Trials of Treatment for Depression." British Medical Journal **303**(6804): 685-692.
Beaumont, G., D. Baldwin, et al. (1996). "A criticism of the practice of prescribing subtherapeutic doses of antidepressants for the treatment of depression." Human Psychopharmacology-Clinical and Experimental **11**(4): 283-291.
Bech, P., K. B. Stage, et al. (1997). "The Major Depression Rating Scale (MDS). Inter-rater reliability and validity across different settings in randomized moclobemide trials." Journal of Affective Disorders **42**(1): 39-48.
Bech, P. and O. J. Rafaelsen (1980). "The Use of Rating-Scales Exemplified by a Comparison of the Hamilton and the Bech-Rafaelsen Melancholia Scale." Acta Psychiatrica Scandinavica **62**: 128-132.
Beck, A., C. Ward, et al. (1961). "An Inventory for Measuring Depression." Arch Gen Psychiatry **4**: 561-571.
Beck, A. T., R. A. Steer, et al. (1988). "Psychometric Properties of the Beck Depression Inventory - 25 Years of Evaluation." Clinical Psychology Review **8**(1): 77-100.
Belz, G., H. Stern, et al. (1983). Cardiovascular effects lofepramine and amitriptyline. VII World Congress of Psychiatry, Vienna.
Beneke, M. and W. Rasmus (1992). "Clinical Global Impressions (ECDEU): some critical comments." Pharmacopsychiatry **25**: 171-176.
Benkert, O., G. Grunder, et al. (1996). "A randomized, double-blind comparison of a rapidly escalating dose of venlafaxine and imipramine in inpatients with major depression and melancholia." Journal of Psychiatric Research **30**(6): 441-451.
Bentkover, J. D. and J. P. Feighner (1995). "Cost-Analysis of Paroxetine Versus Imipramine in Major Depression." Pharmacoeconomics **8**(3): 223-232.
Berger, M., K. Bingefor, et al., Eds. (2003). Health Care Cost, Quality, and Outcomes, ISPOR.
Berndt, E. R., J. M. Russell, et al. (2000). "Comparing SSRI Treatment Costs for Depression Using Retrospective Claims Data: The Role of Nonrandom Selection and Skewed Data." Value in Health **3**(3): 208-221.
Beuzen, J. N., V. F. Ravily, et al. (1993). "Impact of Fluoxetine on Work Loss in

Depression." International Clinical Psychopharmacology **8**(4): 319-321.
Bielski, R. J., D. Ventura, et al. (2004). "A double-blind comparison of escitalopram and venlafaxine extended release in the treatment of major depressive disorder." Journal of Clinical Psychiatry **65**(9): 1190-1196.
Birnbaum, H. (2005). "Friction-cost method as an alternative to the Human-capital approach in calculating indirect costs." Pharmacoeconomics **23**(2): 103-104.
Bisserbe, J. C. (2002). "Clinical utility of milnacipran in comparison with other antidepressants." International Clinical Psychopharmacology **17**: S43-S50.
Blacker, C. V. R. and A. W. Clare (1987). "Depressive Disorder in Primary Care." British Journal of Psychiatry **150**: 737-751.
BNF (2010). British National Formulary. London, British Medical Association & Royal Pharmaceutical Society.
Bondy, B., T. C. Baghai, et al. (2003). "Substance P serum levels are increased in major depression: Preliminary results." Biological Psychiatry **53**(6): 538-542.
Borghi, J. and J. F. Guest (2000). "Economic impact of using mirtazapine compared to amitriptyline and fluoxetine in the treatment of moderate and severe depression in the UK." European Psychiatry **15**(6): 378-387.
Bosc, M., A. Dubini, et al. (1997). "Development and validation of a social functioning scale, the social adaptation self-evaluation scale." European Neuropsychopharmacology **7**: S57-S70.
Bouchard, J. M., J. Delaunay, et al. (1987). "Citalopram Versus Maprotiline - a Controlled, Clinical Multicenter Trial in Depressed-Patients." Acta Psychiatrica Scandinavica **76**(5): 583-592.
Boyer, P., J. M. Danion, et al. (1998). "Clinical and Economic Comparison of Sertraline and Fluoxetine in the Treatment of Depression - a 6-Month Double-Blind Study in a Primary-Care Setting in France." Pharmacoeconomics **13**(1): 157-169.
Brannan, S. K., C. H. Mallinckrodt, et al. (2005). "Onset of action for duloxetine 60 mg once daily: double-blind, placebo-controlled studies." Journal of Psychiatric Research **39**(2): 161-172.
Bremner, J. D. (2002). "Brain pathways in stress processing." Biological Psychiatry **51**(8, Supplement 1): S91-S92.
Brøsen, K. and C. A. Naranjo (2001). "Review of pharmacokinetic and pharmacodynamic interaction studies with citalopram." European Neuropsychopharmacology **11**(4): 275-283.
Brown, M. C. J., J. M. T. Van Loon, et al. (1999a). "Cost-Effectiveness of Mirtazapine Relative to Amitriptyline in the Treatment of Moderate and Severe Depression in France." European Journal of Psychiatry **13**(4): 197-208.
Brown, M. C. J., A. A. Nimmerrichter, et al. (1999b). "Cost-Effectiveness of Mirtazapine Compared to Amitriptyline and Fluoxetine in the Treatment of Moderate and Severe Depression in Austria." European Psychiatry **14**(4): 230-244.
Brown, M. C. J., J. M. T. Van Loon, et al. (2000). "Cost-Effectiveness of Mirtazapine Relative to Fluoxetine in the Treatment of Moderate and Severe Depression in France." European Journal of Psychiatry **14**(1): 15-25.
Brugha, T. S., P. E. Bebbington, et al. (1992). "Antidepressants May Not Assist Recovery in Practice - a Naturalistic Prospective Survey." Acta Psychiatrica Scandinavica **86**(1): 5-11.
Brunello, N. and G. Racagni (1998). "Rationale for the development of noradrenaline reuptake inhibitors." Human Psychopharmacology-Clinical and Experimental

13: S13-S19.
Burke, W. J., I. Gergel, et al. (2002). "Fixed-dose trial of the single isomer SSRI escitalopram in depressed outpatients." Journal of Clinical Psychiatry **63**(4): 331-336.
Burke, W. J. (2004). "Selective versus multi-transmitter antidepressants: Are two mechanisms better than one?" Journal of Clinical Psychiatry **65**: 37-45.
Burt, V. K., M. M. Wohlreich, et al. (2005). "Duloxetine for the treatment of major depressive disorder in women ages 40 to 55 years." Psychosomatics **46**(4): 345-54.
Burton, R., Ed. (2001). The Anatomy of Melancholy, 1651. New York, New York Review of Books.
Bymaster, F. P., L. J. Dreshfield-Ahmad, et al. (2001). "Comparative affinity of duloxetine and venlafaxine for serotonin and norepinephrine transporters in vitro and in vivo, human serotonin receptor subtypes, and other neuronal receptors." Neuropsychopharmacology **25**(6): 871-880.
Campbell, S. and G. MacQueen (2004). "The role of the hippocampus in the pathophysiology of major depression." Journal of Psychiatry & Neuroscience **29**(6): 417-426.
Caro, J. J., M. Salas, et al. (2004). "Modeling the efficiency of reaching a target intermediate end point: A case study in type 2 diabetes in the United States." Value in Health **7**(1): 13-21.
Carroll, B. J., M. Feinberg, et al. (1981). "The Carroll Rating-Scale for Depression .1. Development, Reliability and Validation." British Journal of Psychiatry **138**(MAR): 194-200.
Cassidy, S. and J. Henry (1987). "Fatal Toxicity of Antidepressant Drugs in Overdose." British Medical Journal **295**(6605): 1021-1024.
Charney, D. S. (1998). "Monoamine dysfunction and the pathophysiology and treatment of depression." Journal of Clinical Psychiatry **59**: 11-14.
Chen, F. H., M. B. Larsen, et al. (2005). "The S-enantiomer of R,S-citalopram, increases inhibitor binding to the human serotonin transporter by an allosteric mechanism. Comparison with other serotonin transporter inhibitors." European Neuropsychopharmacology **15**(2): 193-198.
Chisolm D, Healey A, Knapp M (1997). "QALYs and mental health care". Soc Psychiatry Psychiatr Epidemiol 32:68-75.
Citizens' Council (2004). Citizens' Council Report: age. Accessed 5 November 2010.
Cipriani, A., T. A. Furukawa, et al. (2009). "Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis." Lancet **373**(9665): 746-58.
Claghorn, J. L. and M. D. Lesem (1995). "A Double-Blind Placebo-Controlled Study of Org-3770 in Depressed Outpatients." Journal of Affective Disorders **34**(3): 165-171.
Clerc, G. E., P. Ruimy, et al. (1994). "A Double-Blind Comparison of Venlafaxine and Fluoxetine in Patients Hospitalized for Major Depression and Melancholia." International Clinical Psychopharmacology **9**(3): 139-143.
Clerc, G. (2001). "Antidepressant efficacy and tolerability of milnacipran, a dual serotonin and noradrenaline reuptake inhibitor: a comparison with fluvoxamine." International Clinical Psychopharmacology **16**(3): 145-151.
Colonna, L. R., EH. Anderson, HF (2002). Escitalopram is well tolerated and more

efficacious than citalopram in long-term treatment of moderately depressed patients. 3rd International Forum on Mood and Anxiety Disorders, Monte Carlo, Monaco.

Colonna, L., H. F. Andersen, et al. (2005). "A randomized, double-blind, 24-week study of escitalopram (10 mg/day) versus citalopram (20 mg/day) in primary care patients with major depressive disorder." Current Medical Research and Opinion **21**(10): 1659-1668.

Combes, A., G. Peytavin, et al. (2001). "Conduction disturbances associated with venlafaxine." Annals of Internal Medicine **134**(2): 166-167.

Committee on Safety of Medicines (1988). "Committee on Safety of Medicines. Iofepramine (Gamanil) and abnormal blood tests of liver function." Current Problems 23.

Cooper, P. J. and C. G. Fairburn (1986). "The Depressive Symptoms of Bulimia-Nervosa." British Journal of Psychiatry **148**: 268-274.

Cox, J. L., J. M. Holden, et al. (1987). "Detection of Postnatal Depression - Development of the 10-Item Edinburgh Postnatal Depression Scale." British Journal of Psychiatry **150**: 782-786.

Cronbach, L. (1951). "Coefficient alpha and the internal structure of tests." Psychometrika **16**(3): 297-333.

Crown, W. H. (2001). "Antidepressant Selection and Economic Outcome: a Review of Methods and Studies From Clinical Practice." British Journal of Psychiatry **179**: S18-S22.

Crown, W. H., R. L. Obenchain, et al. (1998b). "The application of sample selection models to outcomes research: The case of evaluating the effects of antidepressant therapy on resource utilization." Statistics in Medicine **17**(17): 1943-1958.

Crown, W. H., T. R. Hylan, et al. (1998a). "Antidepressant selection and use and healthcare expenditures - An empirical approach." Pharmacoeconomics **13**(4): 435-448.

Curet, O., G. Damoiseau-Ovens, et al. (1998). "Preclinical profile of befloxatone, a new reversible MAO-A inhibitor." Journal of Affective Disorders **51**(3): 287-303.

Damluji, N. F. and J. M. Ferguson (1988). "Paradoxical Worsening of Depressive Symptomatology Caused by Antidepressants." Journal of Clinical Psychopharmacology **8**(5): 347-349.

Dardennes, R., G. Berdeaux, et al. (1999). "Comparison of the Cost-Effectiveness of Milnacipran (a Snri) With Tcas and Ssris: a Modeling Approach." European Psychiatry **14**(3): 152-162.

Davidson, J., C. D. Turnbull, et al. (1986). "The Montgomery-Asberg Depression Scale: reliability and validity." Acta Psychiatrica Scandinavica. **73**(5):544-8.

Davis, R. and M. I. Wilde (1996). "Mirtazapine. A review of its pharmacology and therapeutic potential in the management of major depression." CNS Drugs **5**(5): 389-402.

deBoer, T. (1996). "The pharmacologic profile of mirtazapine." Journal of Clinical Psychiatry **57**: 19-25.

Delgado PL, Miller HL, Salomon RM, et al (1999). Tryptophan-depletion challenge in depressed patients treated with desipramine or fluoxetine: implications for the role of serotonin in the mechanism of antidepressant action. Biol Psychiatry **46**:212-220

Demyttenaere, K., M. E. A. Hemels, et al. (2005). "A cost-effectiveness model of escitalopram, citalopram, and venlafaxine as first-line treatment for major depressive disorder in Belgium." Clinical Therapeutics **27**(1): 111-124.
Dersimonian, R. and N. Laird (1986). "Metaanalysis in Clinical-Trials." Controlled Clinical Trials **7**(3): 177-188.
Detke, M. J., Y. Lu, et al. (2002a). "Duloxetine, 60 mg once daily, for major depressive disorder: a randomized double-blind placebo-controlled trial." J Clin Psychiatry **63**(4): 308-15.
Detke, M. J., Y. Lu, et al. (2002b). "Duloxetine 60 mg once daily dosing versus placebo in the acute treatment of major depression." J Psychiatr Res **36**(6): 383-90.
DeVane, C. L. (2001). "Substance P: A new era, a new role." Pharmacotherapy **21**(9): 1061-1069.
DeWilde, J., R. Spiers, et al. (1993). "A Double-Blind, Comparative, Multicenter Study Comparing Paroxetine with Fluoxetine in Depressed-Patients." Acta Psychiatrica Scandinavica **87**(2): 141-145.
Dinan, T. G. (1998). "Neuroendocrine markers - Role in the development of antidepressants." Cns Drugs **10**(2): 145-157.
Dinos, S., S. Stevens, et al. (2004). "Stigma: the feelings and experiences of 46 people with mental illness: Qualitative study." Br J Psychiatry **184**(2): 176-181.
Discussion (2004). "In the treatment of depression, does selectivity of action matter?" Journal of Clinical Psychiatry **65 (Suppl 4)**: 60-62.
Donoghue, J. and D. M. Taylor (2000). "Suboptimal use of antidepressants in the treatment of depression." Cns Drugs **13**(5): 365-383.
Dostert, P., M. S. Benedetti, et al. (1997). "Review of the pharmacokinetics and metabolism of reboxetine, a selective noradrenaline reuptake inhibitor." European Neuropsychopharmacology **7**: S23-S35.
Dubini, A., M. Bosc, et al. (1997). "Do noradrenaline and serotonin differentially affect social motivation and behaviour?" European Neuropsychopharmacology **7**: S49-S55.
Duman, R. S., G. R. Heninger, et al. (1997). "A molecular and cellular theory of depression." Archives of General Psychiatry **54**(7): 597-606.
Duman, R. S., J. Malberg, et al. (1999). "Neural plasticity to stress and antidepressant treatment." Biological Psychiatry **46**(9): 1181-1191.
Dunbar, G. C., J. B. Cohn, et al. (1991). "A Comparison of Paroxetine, Imipramine and Placebo in Depressed out-Patients." British Journal of Psychiatry **159**: 394-398.
Dunbar, M., G. Ford, et al. (2000). "A confirmatory factor analysis of the Hospital Anxiety and Depression scale: Comparing empirically and theoretically derived structures." British Journal of Clinical Psychology **39**: 79-94.
Dyer, O. (2003). "Lundbeck broke advertising rules." BMJ **326**(7397): 1004-.
Einarson, T. R., S. Arikian, et al. (1995). "A Model to Evaluate the Cost-Effectiveness of Oral Therapies in the Management of Patients With Major Depressive-Disorders." Clinical Therapeutics **17**(1): 136-153.
Einarson, T. R., A. Addis, et al. (1997). "Pharmacoeconomic Analysis of Venlafaxine in the Treatment of Major Depressive Disorder." Pharmacoeconomics **12**(2): 286-296.
Ekselius, L., L. von Knorring, et al. (1997). "A double-blind multicenter trial comparing sertraline and citalopram in patients with major depression treated in general practice." International Clinical Psychopharmacology **12**(6): 323-331.

Elsborg, L. (1991). "Safety aspects of treatment of the elderly with dementia and/depression. Side effect profile of a specific 5-HT uptake inhibitor." Nordic Journal of Psychiatry **45**(Suppl 23): 29-35.
Enns, M. W., B. J. Cox, et al. (1998). "Confirmatory factor analysis of the Beck Anxiety and Depression Inventories in patients with major depression." Journal of Affective Disorders **47**(1-3): 195-200.
Fairchild, C. J., A. J. Rush, et al. (1986). "Which Depressions Respond to Placebo." Psychiatry Research **18**(3): 217-226.
Faries, D., J. Herrera, et al. (2000). "The responsiveness of the Hamilton Depression Rating Scale." Journal of Psychiatric Research **34**: 3-10.
Fassino, S., G. A. Daga, et al. (2004). "Use of reboxetine in bulimia nervosa: a pilot study." Journal of Psychopharmacology **18**(3): 423-428.
Fawcett, J. and R. L. Barkin (1997). "Efficacy issues with antidepressants." Journal of Clinical Psychiatry **58**: 32-39.
Feighner, J., E. Robins, et al. (1972). "Diagnostic Criteria for Use in Psychiatric Research." Arch Gen Psychiatry **26**: 57-63.
Fernandez, J. L., S. Montgomery, et al. (2005). "Evaluation of the cost effectiveness of escitalopram versus venlafaxine XR in major depressive disorder." Pharmacoeconomics **23**(2): 155-167.
Foglia, J. P., B. G. Pollock, et al. (1997). "Plasma levels of citalopram enantiomers and metabolites in elderly patients." Psychopharmacology Bulletin **33**(1): 109-112.
Forder, J., S. Kavanagh, et al. (1996). "A Comparison of the Cost-Effectiveness of Sertraline Versus Tricyclic Antidepressants in Primary Care." Journal of Affective Disorders **38**(2-3): 97-111.
Francois, C., M. Toumi, et al. (2003). "A pharmacoeconomic evaluation of escitalopram, a new selective serotonin reuptake inhibitor." European Journal of Health Economics **4**: 12-19.
Frank, L., D. A. Revicki, et al. (2001). "The economics of selective serotonin reuptake inhibitors in depression - A critical review." Cns Drugs **15**(1): 59-83.
Freeman, H., S. Arikian, et al. (2000). "Pharmacoeconomic Analysis of Antidepressants for Major Depressive Disorder in the United Kingdom." Pharmacoeconomics **18**(2): 143-148.
Fuller, R. W., K. W. Perry, et al. (1996). "Serum corticosterone increases reflect enhanced uptake inhibitor-induced elevation of extracellular 5-hydroxytryptamine in rat hypothalamus." Journal of Pharmacy and Pharmacology **48**(1): 68-70.
Furukawa, T. A., H. McGuire, et al. (2002). "Meta-analysis of effects and side effects of low dosage tricyclic antidepressants in depression: systematic review." British Medical Journal **325**(7371): 991-995.
Galinowski, A. and P. Lehert (1995). "Structural Validity of MADRS During Antidepressant Treatment." International Clinical Psychopharmacology **10**(3): 157-161.
Garratt, A. M., D. A. Ruta, et al. (1993). "The Sf-36 Health Survey Questionnaire - an Outcome Measure Suitable for Routine Use Within the Nhs." British Medical Journal **306**(6890): 1440-1444.
Gibbons, R. D., D. C. Clark, et al. (1993). "Exactly What Does the Hamilton Depression Rating-Scale Measure." Journal of Psychiatric Research **27**(3): 259-273.
Gilchrist, S. and M. Knapp (1994). "Economics and mental health." Current Opinion in

Psychiatry **7**: 167-172.
Goldstein, D. J., Y. L. Lu, et al. (2004). "Duloxetine in the treatment of depression - A double-blind- placebo-controlled comparison with paroxetine." Journal of Clinical Psychopharmacology **24**(4): 389-399.
Goldstein, D. J., C. Mallinckrodt, et al. (2002). "Duloxetine in the treatment of major depressive disorder: a double-blind clinical trial." J Clin Psychiatry **63**(3): 225-31.
Golkema, Y. (1983). "Heart-effects of lofepramine in depressed patients with and without heart disease." VII World Congress of Psychiatry.
Goodwin, F. K., A. J. Prange, et al. (1982). "Potentiation of Anti-Depressant Effects by L-Triiodothyronine in Tricyclic Non-Responders." American Journal of Psychiatry **139**(1): 34-38.
Gorman, J., A. Korotzer, et al. (2002). "Efficacy comparison of escitalopram and citalopram in the treatment of major depressive disorder: pooled analysis of placebo-controlled trials." CNS Spectrums **7**(2): 40-44.
Greenberg, P. E., L. E. Stiglin, et al. (1993). "The Economic Burden of Depression in 1990." Journal of Clinical Psychiatry **54**(11): 405-418.
Grundy, C. T., K. M. Lunnen, et al. (1994). "The Hamilton Rating-Scale for Depression - One Scale or Many." Clinical Psychology-Science and Practice **1**(2): 197-205.
Guelfi, J. D., C. White, et al. (1995). "Effectiveness of Venlafaxine in Patients Hospitalized for Major Depression and Melancholia." Journal of Clinical Psychiatry **56**(10): 450-458.
Gunnell, D., G. Saperia, et al. (2005). "Selective serotonin reuptake inhibitors (SSR1s) and suicide in adults: meta-analysis of drug company data from placebo controlled, randomised controlled trials submitted to the MHRA's safety review." British Medical Journal **330**(7488): 385-388A.
Gurney, C., T. A. Kerr, et al. (1972). "Studies in Classification of Affective Disorders - Relationship between Anxiety States and Depressive Illnesses .2." British Journal of Psychiatry **121**(561): 162-&.
Guy, W. (1976). "ECDEU Assessment Manual for Psychopharmacology." US NIMH Psychopharmacology Research Branch: 217-222.
Guze, B. H. (1996). "Selective serotonin reuptake inhibitors - Assessment for formulary inclusion." Pharmacoeconomics **9**(5): 430-442.
Hale, A. S. (1994). "The Importance of Accidents in Evaluating the Cost of Ssris - a Review." International Clinical Psychopharmacology **9**(3): 195-201.
Hamilton, M. (1960). "A Rating Scale for Depression." J Neurol Neurosurg Psychiat **23**: 56-62.
Hamilton, M. (1967). "Development of a Rating Scale for Primary Depressive Illness." Brit J soc clin Psychol **6**: 278-296.
Hammond, M. F. (1998). "Rating depression severity in the elderly physically ill patient: Reliability and factor structure of the Hamilton and the Montgomery-Asberg Depression Rating Scales." International Journal of Geriatric Psychiatry **13**(4): 257-261.
Hatziandreu, E. J., R. E. Brown, et al. (1994). "Cost-Utility of Maintenance Treatment of Recurrent Depression With Sertraline Versus Episodic Treatment With Dothiepin." Pharmacoeconomics **5**(3): 249-264.
Healy, D. and C. Whitaker (2003). "Antidepressants and suicide: risk-benefit conundrums." Journal of Psychiatry & Neuroscience **28**(5): 331-337.

Hemels, M. E. H., S. Kasper, et al. (2004b). "Cost-effectiveness analysis of escitalopram: a new SSRI in the first-line treatment of major depressive disorder in Austria." Current Medical Research and Opinion **20**(6): 869-878.
Hemels, M. E. H., S. Kasper, et al. (2004a). "Cost-effectiveness of escitalopram versus citalopram in the treatment of severe depression." Annals of Pharmacotherapy **38**(6): 954-960.
Hemels, M. E., O. Karamustafyoglu, et al. (2005). "Cost-effectiveness analysis of escitalopram in the treatment of major depressive disorder in Turkey." Value in Health **8**(3): 392-392.
Heninger, G. R., P. L. Delgado, et al. (1996). "The revised monoamine theory of depression: A modulatory role for monoamines, based on new findings from monoamine depletion experiments in humans." Pharmacopsychiatry **29**(1): 2-11.
Henry, D. (1992). "Economic analysis as an aid to subsidisation decisions." Pharmacoeconomics **1**(1): 54-67.
Herpfer, I. and K. Lieb (2005). "Substance P receptor antagonists in psychiatry - Rationale for development and therapeutic potential." Cns Drugs **19**(4): 275-293.
Higgins, J. P. T., S. G. Thompson, et al. (2003). "Measuring inconsistency in meta-analyses." BMJ **327**(7414): 557-560.
Higgins, J. P. T. and S. G. Thompson (2002). "Quantifying heterogeneity in a meta-analysis." Statistics in Medicine **21**(11): 1539-1558.
Holm, K. J. and C. Spencer (1999). "Reboxetine. A review of its use in depression." CNS Drugs **12**(1): 65-83.
Holm, K. J., B. Jarvis, et al. (2000). "Mirtazapine - a Pharmacoeconomic Review of Its Use in Depression." Pharmacoeconomics **17**(5): 515-534.
Horst, W. D. and S. H. Preskorn (1998). "Mechanisms of action and clinical characteristics of three atypical antidepressants: venlafaxine, nefazodone, bupropion." Journal of Affective Disorders **51**(3): 237-254.
Hudson, J., M. Wolreich, et al. (2004). Safetry and tolerability of duloxetine: analysis of pooled data from eight acute phase placebo-controlled clinical trials. European College of Neuropsychopharmacology, Stockholm, Sweden.
Hylan, T. R., W. H. Crown, et al. (1998). "Tricyclic antidepressant and selective serotonin reuptake inhibitors antidepressant selection and health care costs in the naturalistic setting: a multivariate analysis." Journal of Affective Disorders **47**(1-3): 71-79.
Hyttel, J. (1994). "Pharmacological Characterization of Selective Serotonin Reuptake Inhibitors (Ssris)." International Clinical Psychopharmacology **9**: 19-26.
Jeffery, K. J. and I. C. Reid (1997). "Modifiable neuronal connections: An overview for psychiatrists." American Journal of Psychiatry **154**(2): 156-164.
Jenkinson, C., S. Stewart-Brown, et al. (1999). "Assessment of the Sf-36 Version 2 in the United Kingdom." Journal of Epidemiology and Community Health **53**(1): 46-50.
Jiang, W., J. Alexander, et al. (2001). "Relationship of Depression to Increased Risk of Mortality and Rehospitalization in Patients With Congestive Heart Failure." Arch Intern Med **161**(15): 1849-1856.
Johansson, C., M. Jansson, et al. (2001). "Genetics of affective disorders." European Neuropsychopharmacology **11**(6): 385-394.

Johnson, D. A. W. (1974). "Study of Use of Antidepressant Medication in General-Practice." British Journal of Psychiatry **125**(AUG): 186-192.
Jonsson, B. and P. E. Bebbington (1993). Economic studies of the treatment of depressive illness. Health Economics of Depression. B. Jonsson and J. Rosenbaum. Chichester, Wiley. **4**.
Jonsson, B. and P. E. Bebbington (1994). "What Price Depression - the Cost of Depression and the Cost-Effectiveness of Pharmacological Treatment." British Journal of Psychiatry **164**: 665-673.
Judd, L. (1995). "Mood disorders in the general population represent an important and worldwide public health problem." International Clinical Psychopharmacology **10**(Suppl 4): 5-10.
Kamlet, M. S., N. Paul, et al. (1995). "Cost-Utility Analysis of Maintenance Treatment for Recurrent Depression." Controlled Clinical Trials **16**(1): 17-40.
Kasper, S. (1995). "Clinical efficacy of mirtazapine: A review of meta-analyses of pooled data." International Clinical Psychopharmacology **10**: 25-35.
Kasper, S., Y. Pletan, et al. (1996). "Comparative studies with milnacipran and tricyclic antidepressants in the treatment of patients with major depression: A summary of clinical trial results." International Clinical Psychopharmacology **11**: 35-39.
Kasper S, Praschak-Reider N, Tauscher J, Wolf R 1997. A risk-benefit assessment of mirtazapine in the treatment of depression. Drug Safety **17**(4):251-264.
Keck Jr., P. E., J. A. Welge, et al. (2000). "Placebo effect in randomized, controlled studies of acute bipolar mania and depression." Biological Psychiatry **47**(8): 748-755.
Kennedy, S. H., H. F. Andersen, et al. (2006). "Efficacy of escitalopram in the treatment of major depressive disorder compared with conventional selective serotonin reuptake inhibitors and venlafaxine XR: a meta-analysis." J Psychiatry Neurosci **31**(2): 122-31.
Kind, P. and J. Sorensen (1993). "The Costs of Depression." International Clinical Psychopharmacology **7**(3-4): 191-195.
Kind, P. and J. Sorensen (1995). "Modeling the Cost-Effectiveness of the Prophylactic Use of Ssris in the Treatment of Depression." International Clinical Psychopharmacology **10**: 41-48.
Kirsch, I. (2000). "Are drug and placebo effects in depression additive?" Biological Psychiatry **47**(8): 733-735.
Kitayama, I., T. Yaga, et al. (1997). "Long-term stress degenerates, but imipramine regenerates, noradrenergic axons in the rat cerebral cortex." Biological Psychiatry **42**(8): 687-696.
Kroenke, K., S. L. West, et al. (2001). "Similar effectiveness of paroxetine, fluoxetine, and sertraline in primary care - A randomized trial." Jama-Journal of the American Medical Association **286**(23): 2947-2955.
Kromer, W. (2001). "Relative efficacies of gastric proton-pump inhibitors on a milligram basis: Desired and undesired SH reactions. Impact of chirality." Scandinavian Journal of Gastroenterology **36**: 3-9.
Kulp, W., W. Greiner, et al. (2005). "Cost-effectiveness of Escitalopram vs. Venlafaxin XR in the treatment of major depressive disorder in Germany." Value in Health **8**(3): 392-393.
Kupfer, D. J. (1991). "Long-Term Treatment of Depression." Journal of Clinical Psychiatry **52**: 28-34.

Lambert, M., D. Hatch, et al. (1986). "Zung, Beck, and Hamilton rating scales as measures of treatment outcome: a meta-analytic comparison." J Consulting and Clin Psychology **54**(1): 54-59.
Lambert, M. J., K. S. Masters, et al. (1988). "An Effect-Size Comparison of the Beck, Zung, and Hamilton Rating-Scales for Depression - a 3-Week and 12-Week Analysis." Psychological Reports **63**(2): 467-470.
Lancaster, S. G. and J. P. Gonzalez (1989). "Lofepramine - a Review of Its Pharmacodynamic and Pharmacokinetic Properties, and Therapeutic Efficacy in Depressive-Illness." Drugs **37**(2): 123-140.
Lane, R. M. and G. B. Baker (1999). "Chirality and drugs used in psychiatry: Nice to know or need to know?" Cellular and Molecular Neurobiology **19**(3): 355-372.
Lapierre, Y. D. (2003). "Suicidality with selective serotonin reuptake inhibitors: Valid claim?" Journal of Psychiatry & Neuroscience **28**(5): 340-347.
Lapierre, Y., J. Bentkover, et al. (1995). "Direct Cost of Depression - Analysis of Treatment Costs of Paroxetine Versus Imipramine in Canada." Canadian Journal of Psychiatry-Revue Canadienne De Psychiatrie **40**(7): 370-377.
Lavori, P. W. (2000). "Placebo control groups in randomized treatment trials: a statistician's perspective." Biological Psychiatry **47**(8): 717-723.
Learned-Coughlin, S. M., M. Bergstrom, et al. (2003). "In vivo activity of bupropion at the human dopamine transporter as measured by positron emission tomography." Biological Psychiatry **54**(8): 800-805.
Lejoyeux, M. and J. Ades (1997). "Antidepressant discontinuation: A review of the literature." Journal of Clinical Psychiatry **58**: 11-16.
Lenox-Smith, A., P. Conway, et al. (2004). "Cost effectiveness of representatives of three classes of antidepressants used in major depression in the UK." Pharmacoeconomics **22**(5): 311-319.
Leon, A. C., M. K. Shear, et al. (1993). "Effect size as a measure of symptom-specific drug change in clinical trials." Psychopharmacology Bulletin **29**(2): 163-7.
Leonard, B. E. (1996). The comparative pharmacological properties of selective serotonin reuptake inhibitors in animals. Selective Serotonin Reuptake Inhibitors. Advances in Basic Research and Clinical Practice. J. P. Feighner and W. Boyer. Chichester, England, John Wiley & Sons Ltd.
Lepen, C., E. Levy, et al. (1994). "The Cost of Treatment Dropout in Depression - a Cost-Benefit- Analysis of Fluoxetine Vs Tricyclics." Journal of Affective Disorders **31**(1): 1-18.
Lepine, J. P., M. Gastpar, et al. (1997). "Depression in the community: The first pan-European study DEPRES (Depression Research in European Society)." International Clinical Psychopharmacology **12**(1): 19-29.
Lepola, U. M., H. Loft, et al. (2003). "Escitalopram (10-20 mg/day) is effective and well tolerated in a placebo-controlled study in depression in primary care." International Clinical Psychopharmacology **18**(4): 211-217.
Lepola, U., A. Wade, et al. (2004). "Do equivalent doses of escitalopram and citalopram have similar efficacy? A pooled analysis of two positive placebo-controlled studies in major depressive disorder." International Clinical Psychopharmacology **19**(3): 149-155.
Levinson, D. F. (2005). "The Genetics of Depression: A Review." Biological Psychiatry (In Press, Corrected Proof). Accessed December 2005.
Li Wan Po, A. (1999). "Antidepressants and upper gastrointestinal bleeding - New

results suggest a link." British Medical Journal **319**(7217): 1081-1082.
Licht, R. W., S. Qvitzau, et al. (2005). "Validation of the Bech-Rafaelsen Melancholia Scale and the Hamilton Depression Scale in patients with major depression; is the total score a valid measure of illness severity?" Acta Psychiatrica Scandinavica **111**(2): 144-149.
Llorca P-M, Azorin J-M, Despiegel N, Verpillat P (2005). Efficacy of escitalopram in patients with severe depression: a pooled analysis. Int J Clin Pract **59**(3):268-275.
Lothgren, M., M. Hemels, et al. (2004). "A cost-effectiveness analysis of escitalopram as first line treatment of depression in Sweden." Primary Care Psychiatry **9**(4): 153-161.
Lovibond, P. F. and S. H. Lovibond (1995). "The Structure of Negative Emotional States - Comparison of the Depression Anxiety Stress Scales (Dass) with the Beck Depression and Anxiety Inventories." Behaviour Research and Therapy **33**(3): 335-343.
Maas, J. W. (1975). "Biogenic-Amines and Depression - Biochemical and Pharmacological Separation of 2 Types of Depression." Archives of General Psychiatry **32**(11): 1357-1361.
Maier, W. and M. Philipp (1985). "Comparative analysis of observer depression scales." Acta psychitr scand **72**: 239-245.
Malberg, J. E., A. J. Eisch, et al. (2000). "Chronic antidepressant treatment increases neurogenesis in adult rat hippocampus." Journal of Neuroscience **20**(24): 9104-9110.
Mallinckrodt, C. H., W. S. Clark, et al. (2001). "Accounting for dropout bias using mixed-effects models." J Biopharm Stat **11**(1-2): 9-21.
Mallinckrodt, C. H., T. M. Sanger, et al. (2003). "Assessing and interpreting treatment effects in longitudinal clinical trials with missing data." Biol Psychiatry **53**(8): 754-60.
Mann R, Gilbody S, Richards D (2009). Putting the 'Q' in depression QALYs: a comparison of utility measurement using EQ-5D and SF-6D health related quality of life measures. Soc Psychiatry Psychiatr Epidemiol **44**:569–578.
Masand, P. S. (2003). "Tolerability and adherence issues in antidepressant therapy." Clinical Therapeutics **25**(8): 2289-2304.
Mattocks, K. M. and R. I. Horwitz (2000). "Placebos, active control groups, and the unpredictability paradox." Biological Psychiatry **47**(8): 693-698.
Mendels, J., F. Reimherr, et al. (1995). "A Double-Blind, Placebo-Controlled Trial of 2 Dose Ranges of Nefazodone in the Treatment of Depressed Outpatients." Journal of Clinical Psychiatry **56**: 30-36.
Mendlewicz, J. (1995). "Pharmacological Profile and Efficacy of Venlafaxine." International Clinical Psychopharmacology **10**: 5-13.
Miller, H. L., P. L. Delgado, et al. (1996a). "Effects of alpha-methyl-para-tyrosine (AMPT) in drug-free depressed patients." Neuropsychopharmacology **14**(3): 151-157.
Miller, H. L., P. L. Delgado, et al. (1996b). "Clinical and biochemical effects of catecholamine depletion on antidepressant-induced remission of depression." Archives of General Psychiatry **53**(2): 117-128.
Moher, D., K. F. Schulz, et al. (2001). "The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group

randomised trials." Lancet **357**(9263): 1191-1194.
Moncrieff, J. and I. Kirsch (2005). "Efficacy of antidepressants in adults." British Medical Journal **331**(7509): 155-157.
Montgomery, S. A. (1994). "Clinically relevant effect sizes in depression." European Neuropsychopharmacology **4**(3): 283-284.
Montgomery, S. A. (1995). "Rapid Onset of Action of Venlafaxine." International Clinical Psychopharmacology **10**: 21-27.
Montgomery, S. A. (1997a). "Is there a role for a pure noradrenergic drug in the treatment of depression?" European Neuropsychopharmacology **7**: S3-S9.
Montgomery, S. A. (1997b). "Reboxetine: additional benefits to the depressed patient." Journal of Psychopharmacology **11**(4 Suppl): S9-S15.
Montgomery, S. A. (1999a). "Alternatives to placebo-controlled trials in psychiatry. ECNP Consensus Meeting, September 26, 1996, Amsterdam. European College of Neuropsychopharmacology." Eur Neuropsychopharmacol **9**(3): 265-9.
Montgomery, S. A. (1999b). "The failure of placebo-controlled studies." European Neuropsychopharmacology **9**(3): 271-276.
Montgomery, S. A. and M. Åsberg (1979). "A New Depression Scale Designed to be Sensitive to Change." Brit J Psychiat **134**: 382-389.
Montgomery, S. A. and R. M. Pinder (1987). "Do Some Antidepressants Promote Suicide." Psychopharmacology **92**(2): 265-266.
Montgomery SA, Henry J, McDonald G, Dinan T et al (1994). Selective serotonin reuptake inhibitors: meta-analysis of discontinuation rates. Int Clin Psychopharmacol **9**:47-53.
Montgomery, S. A., H. Loft, et al. (2001). "Escitalopram (S-Enantiomer of citalopram): Clinical efficacy and onset of action predicted from a rat model." Pharmacology & Toxicology **88**(5): 282-286.
Montgomery, S. A., A. K. T. Huusom, et al. (2004). "A randomised study comparing escitalopram with venlafaxine XR in primary care patients with major depressive disorder." Neuropsychobiology **50**(1): 57-64.
Moore, N., H. Verdoux, et al. (2005). "Prospective, multicentre, randomized, double-blind study of the efficacy of escitalopram versus citalopram in outpatient treatment of major depressive disorder." International Clinical Psychopharmacology **20**(3): 131-137.
Moran, P. and M. J. Lambert (1983). A review of current assessment tools for monitoring changes in depression. New York, Wiley.
Mork, A. and C. Sanchez (2003a). "The R-enantiomer of citalopram counteracts escitalopram-induced in vivo effects in rats." European Neuropsychopharmacology **13**: S204-S204.
Mork, A., M. Kreilgaard, et al. (2003b). "The R-enantiomer of citalopram increase in extracellular 5-HT counteracts escitalopram-induced in the frontal cortex of freely moving rats." Neuropharmacology **45**(2): 167-173.
Muller-Oerlinghausen, B. and A. Berghofer (1999). "Antidepressants and suicidal risk." J Clin Psychiatry **60**(Suppl 2): 94-99.
Murray, C. J. L. and A. D. Lopez (1997). "Alternative projections of mortality and disability by cause 1990-2020: Global burden of disease study." Lancet **349**(9064): 1498-1504.
Musselman, D. L., C. DeBattista, et al. (1998). Chapter 27 Biology of Mood Disorders.

Textbook of Psychopharmacology. S. A. F. and C. B. Nemeroff. Washington, APA.
Nelson, J. C., L. Portera, et al. (2005). "Are There Differences in the Symptoms that Respond to a Selective Serotonin or Norepinephrine Reuptake Inhibitor?" Biological Psychiatry **57**(12): 1535-1542.
Nestler, E. J., R. Z. Terwilliger, et al. (1989). "Chronic Antidepressant Administration Alters the Subcellular- Distribution of Cyclic Amp-Dependent Protein-Kinase in Rat Frontal-Cortex." Journal of Neurochemistry **53**(5): 1644-1647.
Nestler, E. J. (1998). "Antidepressant treatments in the 21st century." Biological Psychiatry **44**(7): 526-533.
Neumeister, A., S. Wood, et al. (2005). "Reduced hippocampal volume in unmedicated, remitted patients with major depression versus control subjects." Biological Psychiatry **57**(8): 935-937.
Nibuya, M., E. J. Nestler, et al. (1996). "Chronic antidepressant administration increases the expression of cAMP response element binding protein (CREB) in rat hippocampus." Journal of Neuroscience **16**(7): 2365-2372.
NICE (2004). Depression: management of depression in primary and secondary care. Clinical practice guideline No 23. London, National Institute for Health and Clinical Excellence.
NICE (2002). Schizophrenia: core interventions in the treatment and management of schizophrenia in primary and secondary care. Clinical Practice Guideline No 1. London, National Institute for Health and Clinical Excellence.
NICE (2007). NICE complies with judge's ruling on judicial review. http://www.nice.org.uk/newsevents/infocus/infocusarchive/NICECompliesWithJudgesRulingOnJudicialReview.jsp. Accessed 4Feb2008.
NICE (2009). Depression.The treatment and management of depression in adults. Clinical Guideline 90.
Nierenberg, A. A., J. P. Feighner, et al. (1994). "Venlafaxine for Treatment-Resistant Unipolar Depression." Journal of Clinical Psychopharmacology **14**(6): 419-423.
Noble, J., John H. (2006). "Meta-analysis: Methods, strengths, weaknesses, and political uses." Journal of Laboratory and Clinical Medicine **147**(1): 7-20.
Nuijten, M. J. C., M. Hardens, et al. (1995). "A Markov Process Analysis Comparing the Cost-Effectiveness of Maintenance Therapy With Citalopram Versus Standard Therapy in Major Depression." Pharmacoeconomics **8**(2): 159-168.
Nuijten, M., L. Hadjadjeba, et al. (1998). "Cost Effectiveness of Fluvoxamine in the Treatment of Recurrent Depression in France." Pharmacoeconomics **14**(4): 433-445.
Nurnberg, H. G., P. M. Thompson, et al. (1999). "Antidepressant medication change in a clinical treatment setting: A comparison of the effectiveness of selective serotonin reuptake inhibitors." Journal of Clinical Psychiatry **60**(9): 574-579.
Nutt, D. J. (2002). "The neuropharmacology of serotonin and noradrenaline in depression." International Clinical Psychopharmacology **17**: S1-S12.
Osman, A., W. Downs, et al. (1997). "Factor structure and psychometric characteristics of the Beck Depression Inventory-II." J Psychopathology and Behavioral Assessment **19**(4): 359-376.
Ottosson, J. O. (1960). "Experimental studies of the mode of action of electroconvulsive therapy: Introduction." Acta Psychiatr Scand **35(Suppl 145)**: 5-6.

Owens, M. J. and C. B. Nemeroff (1994). "Role of Serotonin in the Pathophysiology of Depression - Focus on the Serotonin Transporter." Clinical Chemistry **40**(2): 288-295.

Owens, M. J., D. L. Knight, et al. (2001). "Second-generation SSRIs: Human monoamine transporter binding profile of escitalopram and R-fluoxetine." Biological Psychiatry **50**(5): 345-350.

Palmer, C. S., L. Kleinman, et al. (1998). "Pharmacoeconomics of antidepressant drug overdose." Cns Drugs **10**(3): 223-231.

Pande, A. C. and M. E. Sayler (1993). "Adverse Events and Treatment Discontinuations in Fluoxetine Clinical-Trials." International Clinical Psychopharmacology **8**(4): 267-269.

Parker, R. D., E. P. Flint, et al. (2003). "A three-factor analytic model of the MADRS in geriatric depression." International Journal of Geriatric Psychiatry **18**(1): 73-77.

Patris, M., J. M. Bouchard, et al. (1996). "Citalopram versus fluoxetine: A double-blind, controlled, multicentre, phase III trial in patients with unipolar major depression treated in general practice." International Clinical Psychopharmacology **11**(2): 129-136.

Peveler, R., T. Kendrick, et al. (2005). A randomised controlled trial to compare the cost-effectiveness of tricyclic antidepressants, selective serotonin reuptake inhibitors and lofepramine. Health Technology Assessment. **2005**.

Pirraglia, P. A., A. B. Rosen, et al. (2004). "Cost-utility analysis studies of depression management: A systematic review." American Journal of Psychiatry **161**(12): 2155-2162.

Plenge, P. and E. T. Mellerup (1985). "Antidepressive drugs can change the affinity of [^{3}H]imipramine and [^{3}H]paroxetine binding to platelet and neuronal membranes." European Journal of Pharmacology **119**: 1-8.

Plenge, P., E. T. Mellerup, et al. (1991). "Affinity modulation of [^{3}H]imipramine, [^{3}H]paroxetine and [^{3}H]citalopram binding to the 5-HT transporter from brain and platelets." European Journal of Pharmacology **206**: 243-250.

Pop, V. J., I. H. Komproe, et al. (1992). "Characteristics of the Edinburgh Post Natal Depression Scale in the Netherlands." Journal of Affective Disorders **26**(2): 105-110.

Posternak, M. A. and M. Zimmerman (2005). "Is there a delay in the antidepressant effect? A meta-analysis." Journal of Clinical Psychiatry **66**(2): 148-158.

PPA (2001). Prescribing of mental health drugs. PPA: News - Pact Pages, Prescription Pricing Authority. **2005**.

Priest, R. G. (1996). "Cost-Effectiveness of Venlafaxine for the Treatment of Major Depression in Hospitalized Patients." Clinical Therapeutics **18**(2): 347-358.

Procopio, M. (2005). "The multiple outcomes bias in antidepressants research." Medical Hypotheses **65**(2): 395-399.

Prozac (2005). Specification of Product Characteristics, emc.medicines. **2005**.

Puech, A., S. A. Montgomery, et al. (1997). "Milnacipran, a new serotonin and noradrenaline reuptake inhibitor: An overview of its antidepressant activity and clinical tolerability." International Clinical Psychopharmacology **12**(2): 99-108.

Pugh, R., J. Bell, et al. (1982). "Does Lofepramine Have Fewer Side-Effects Than Amitriptyline - Results of a Comparative Trial." Journal of Affective Disorders **4**(4): 355-363.

Rabkin, J. G., J. S. Markowitz, et al. (1986). "How blind is blind? Assessment of patient

and doctor medication guesses in a placebo-controlled trial of imipramine and phenelzine." Psychiatry Research **19**(1): 75-86.
Ramklint, M. and L. Ekselius (2003). "Personality traits and personality disorders in early onset versus late onset major depression." Journal of Affective Disorders **75**(1): 35-42.
Ramos-Brieva, J. A. and A. Cordero-Villafafila (1988). "A new validation of the Hamilton Rating Scale for Depression." Journal of Psychiatric Research **22**(1): 21-8.
Rampello, L., A. Alvano, et al. (2005). "An evaluation of efficacy and safety of reboxetine in elderly patients affected by "retarded" post-stroke depression: A random, placebo-controlled study." Archives of Gerontology and Geriatrics **40**(3): 275-285.
Rapaport, M. H., A. Bose, et al. (2004). "Escitalopram continuation treatment prevents relapse of depressive episodes." Journal of Clinical Psychiatry **65**(1): 44-49.
Reines, E. H., H. Loft, et al. (2002). "Escitalopram is efficacious and well tolerated in the treatment of depression in primary care." European Neuropsychopharmacology **12**: S254-S254.
Reines, E. H. and N. Despiegel (2002). Switching depressed patients from citalopram to escitalopram treatment is safe and effective. 3rd International Forum on Mood and Anxiety Disorders, Monte Carlo, Monaco.
Revicki, D. A., R. E. Brown, et al. (1995). "Modeling the Cost-Effectiveness of Antidepressant Treatment in Primary-Care." Pharmacoeconomics **8**(6): 524-540.
Revicki, D. A., R. E. Brown, et al. (1997a). "Cost-Effectiveness of Newer Antidepressants Compared With Tricyclic Antidepressants in Managed Care Settings." Journal of Clinical Psychiatry **58**(2): 47-58.
Revicki, D. A., C. S. Palmer, et al. (1997b). "Acute Medical Costs of Fluoxetine Versus Tricyclic Antidepressants - a Prospective Multicentre Study of Antidepressant Drug Overdoses." Pharmacoeconomics **11**(1): 48-55.
Rich, C. L. (1999). "Relationship between antidepressant treatment and suicide [letter]." J Clin Psychiatry **60**(5): 340.
Richelson, E. (1979). "Tricyclic Anti-Depressants and Histamine-H1 Receptors." Mayo Clinic Proceedings **54**(10): 669-674.
Richelson, E. (1996). "Synaptic effects of antidepressants." Journal of Clinical Psychopharmacology **16**(3): S1-S7.
Richou, H., P. Ruimy, et al. (1995). "A Multicenter, Double-Blind, Clomipramine-Controlled Efficacy and Safety Study of Org-3770." Human Psychopharmacology-Clinical and Experimental **10**(4): 263-271.
Rickels, K., A. Derivan, et al. (1995). "Rapid onset of antidepressant activity with venlafaxine treatment." Depression **3**: 146-153.
Rocca, P., V. Fonzo, et al. (2002). "A comparison of paroxetine and amisulpride in the treatment of dysthymic disorder." Journal of Affective Disorders **70**(3): 313-317.
Roeloffs, C., C. Sherbourne, et al. (2003). "Stigma and depression among primary care patients." General Hospital Psychiatry **25**(5): 311-315.
Romeo, R., A. Patel, et al. (2004). "The cost-effectiveness of mirtazapine versus paroxetine in treating people with depression in primary care." International Clinical Psychopharmacology **19**(3): 125-134.
Rosenberg, C., N. Damsbo, et al. (1994). "Citalopram and Imipramine in the Treatment of Depressive Patients in General-Practice - a Nordic Multicenter Clinical-

Study." International Clinical Psychopharmacology **9**: 41-48.
Rosenthal, M. H. and D. Li (2002, October 5-9). Efficacy and tolerability of escitalopram in patients intolerant of other SSRIs. 15th Congress of the European College of Neuropsychopharmacology, Barcelona, Spain.
Rush, A. J., C. M. Gullion, et al. (1996). "The Inventory of Depressive Symptomatology (IDS): Psychometric properties." Psychological Medicine **26**(3): 477-486.
Sacristan, J. A., I. Gilaberte, et al. (2000). "Cost-Effectiveness of Fluoxetine Plus Pindolol in Patients With Major Depressive Disorder: Results From a Randomized, Double- Blind Clinical Trial." International Clinical Psychopharmacology **15**(2): 107-113.
Salter, M. (1996). "Selective inhibitors of tryptophan 2,3 dioxygenase and combined inhibitors of tryptophan 2,3 dioxygenase and 5-HT reuptake as novel serotonergic antidepressants." CNS Drug Reviews **2**(1): 127-143.
Sanchez, C., E. Bien, et al. (2003a). "R-citalopram counteracts the effect of escitalopram in a rat conditioned fear stress model of anxiety." European Neuropsychopharmacology **13**: S211-S211.
Sanchez, C., P. B. F. Bergqvist, et al. (2003b). "Escitalopram, the S-(+)-enantiomer of citalopram, is a selective serotonin reuptake inhibitor with potent effects in animal models predictive of antidepressant and anxiolytic activities." Psychopharmacology **167**(4): 353-362.
Sanchez, C. (2003). "R-citalopram attenuates anxiolytic effects of escitalopram in a rat ultrasonic vocalisation model." European Journal of Pharmacology **464**(2-3): 155-158.
Sanchez, C., K. P. Bogeso, et al. (2004a). "Escitalopram versus citalopram: the surprising role of the R- enantiomer." Psychopharmacology **174**(2): 163-176.
Sanchez, C. and M. Kreilgaard (2004). "R-citalopram inhibits functional and 5-HTP-evoked behavioural responses to the SSRI, escitalopram." Pharmacology Biochemistry and Behavior **77**(2): 391-398.
Schaaf, M. J. M., J. de Jong, et al. (1998). "Downregulation of BDNF mRNA and protein in the rat hippocampus by corticosterone." Brain Research **813**(1): 112-120.
Schatzberg, A. F. and H. C. Kraemer (2000). "Use of placebo control groups in evaluating efficacy of treatment of unipolar major depression." Biological Psychiatry **47**(8): 736-744.
Schildkraut, J. (1969). "Neuropsychopharmacology and the affective disorders (third of three parts)." New Eng J Med **281**: 302-308.
Sclar, D. A., L. M. Robison, et al. (1995). "Antidepressant pharmacotherapy: Economic evaluation of fluoxetine, paroxetine and sertraline in a health maintenance organization." Journal of International Medical Research **23**(6): 395-412.
Sclar, D. A., T. L. Skaer, et al. (1999). "Economic appraisal of citalopram in the management of single-episode depression." Journal of Clinical Psychopharmacology **19**(5): 47S-54S.
Senra, C. (1995). "Measures of treatment outcome of depression: an effect-size comparison." Psychological Reports **76**(1): 187-92.
Shaw, D. M., D. R. Thomas, et al. (1986). "A Comparison of the Antidepressant Action of Citalopram and Amitriptyline." British Journal of Psychiatry **149**: 515-517.
Sheline, Y. I., P. W. Wang, et al. (1996). "Hippocampal atrophy in recurrent major depression." Proceedings of the National Academy of Sciences of the United States of America **93**(9): 3908-3913.

Sheline, Y. I., M. H. Gado, et al. (2003). "Untreated Depression and Hippocampal Volume Loss." Am J Psychiatry **160**(8): 1516-1518.
Sidhu, J., M. Priskorn, et al. (1997). "Steady-state pharmacokinetics of the enantiomers of citalopram and its metabolites in humans." Chirality **9**(7): 686-692.
Simon, G., E. Wagner, et al. (1995). "Cost-Effectiveness Comparisons Using Real-World Randomized Trials - the Case of New Antidepressant Drugs." Journal of Clinical Epidemiology **48**(3): 363-373.
Simon, G. E., M. VonKorff, et al. (1996). "Initial antidepressant choice in primary care - Effectiveness and cost of fluoxetine vs tricyclic antidepressants." Jama-Journal of the American Medical Association **275**(24): 1897-1902.
Simon, G. E., J. Heiligenstein, et al. (1999). "Long-term outcomes of initial antidepressant drug choice in a "real world" randomized trial." Archives of Family Medicine **8**(4): 319-325.
Smith, M. A., S. Makino, et al. (1995). "Stress and Glucocorticoids Affect the Expression of Brain- Derived Neurotrophic Factor and Neurotrophin-3 Messenger-Rnas in the Hippocampus." Journal of Neuroscience **15**(3): 1768-1777.
Snaith, R. P., G. W. K. Bridge, et al. (1976). "Leeds Scales for Self-Assessment of Anxiety and Depression." British Journal of Psychiatry **128**(FEB): 156-165.
Snaith, P. (1993). "What Do Depression Rating-Scales Measure." British Journal of Psychiatry **163**: 293-298.
Solomon, D. A., M. B. Keller, et al. (2000). "Multiple recurrences of major depressive disorder." American Journal of Psychiatry **157**(2): 229-233.
Sorbera, L. A., L. Revel, et al. (2001). "Escitalopram oxalate - Antidepressant - 5-HT reuptake inhibitor." Drugs of the Future **26**(2): 115-120.
Spitzer, R., J. Endicott, et al. (1978). "Research Diagnostic Criteria." Arch Gen Psychiatry **35**: 773-782.
Stahl, S. M. (1998). "Mechanism of action of serotonin selective reuptake inhibitors: Serotonin receptors and pathways mediate therapeutic effects and side effects." Journal of Affective Disorders **51**(3): 215-235.
Stahl, S. M. (2000b). "Blue genes and the mechanism of action of antidepressants." Journal of Clinical Psychiatry **61**(3): 164-165.
Stahl, S. M. (2000a). "Blue genes and the monoamine hypothesis of depression." Journal of Clinical Psychiatry **61**(2): 77-78.
Stahl, S. M. (2000c). Essential Psychopharmacology. Neuroscientific Basis and Practical Applications. 2nd edn. Cambridge, UK, Cambridge University Press.
Starkman, M. N., S. S. Gebarski, et al. (1992). "Hippocampal-Formation Volume, Memory Dysfunction, and Cortisol- Levels in Patients with Cushings-Syndrome." Biological Psychiatry **32**(9): 756-765.
Stewart, A. (1994). "Antidepressant pharmacotherapy: cost comparison of SSRIs and TCAs." Brit J Med Economics **7**: 67-79.
Stoudemire, A., R. Frank, et al. (1986). "The Economic Burden of Depression." General Hospital Psychiatry **8**(6): 387-394.
Streiner, D. L. (1993). "A Checklist for Evaluating the Usefulness of Rating-Scales." Canadian Journal of Psychiatry-Revue Canadienne De Psychiatrie **38**(2): 140-148.
Sturm, R., J. Unutzer, et al. (1999). "Effectiveness research and implications for study design: sample size and statistical power." General Hospital Psychiatry **21**(4):

274-283.
Sullivan, P. W., R. Valuck, et al. (2004). "A comparison of the direct costs and cost effectiveness of serotonin reuptake inhibitors and associated adverse drug reactions." Cns Drugs **18**(13): 911-932.
Suter, K. L. and A. Biddle (2003). "The cost-utility of bupropion SR versus sertraline in the treatment of late-life depression." Value in Health **6**(3): 358-359.
Svensson, S. and P. R. Mansfield (2004). "Escitalopram: Superior to citalopram or a chiral chimera?" Psychotherapy and Psychosomatics **73**(1): 10-16.
Szerman, N., L. Peris, et al. (2005). "Reboxetine for the treatment of patients with Cocaine Dependence Disorder." Human Psychopharmacology-Clinical and Experimental **20**(3): 189-192.
Tajima, O. (2002). "Japanese experience with dual-action antidepressants." International Clinical Psychopharmacology **17**: S37-S42.
Taylor, D. P., R. B. Carter, et al. (1995). "Pharmacology and Neurochemistry of Nefazodone, a Novel Antidepressant Drug." Journal of Clinical Psychiatry **56**: 3-11.
Taylor, S. and M. B. Stein (2006). "The future of selective serotonin reuptake inhibitors (SSRIs) in psychiatric treatment." Medical Hypotheses **66**(1): 14-21.
Thase, M. E. (2000). "Treatment of severe depression." Journal of Clinical Psychiatry **61**: 17-25.
Thayer, L. A., V. G. Scott, et al. (2003). "Evaluating the cost-effectiveness of St. John's Wort versus fluoxetine for the treatment of mild to moderate depression." Value in Health **6**(3): 357-357.
Thomas, C. M. and S. Morris (2003). "Cost of depression among adults in England in 2000." British Journal of Psychiatry **183**: 514-519.
Thompson, S. G. and J. P. Higgins (2005). "Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit?" Lancet **365**(9456): 341-6.
Tome, M. B. and M. T. Isaac (1997). "Cost-Benefit & Cost-Effectiveness Analysis of the Rapid Onset of Selective Selective Reuptake Inhibitors by Augmentation." International Journal of Psychiatry in Medicine **27**(4): 377-390.
Tome, M. B. and M. T. Isaac (1998). "Cost Effectiveness Study of a Year Follow-up of Selective Serotonin Reuptake Inhibitor(Ssri) and Augmentor Combination Compared With Ssri and Placebo." International Clinical Psychopharmacology **13**(4): 175-182.
Tucker, G. T. (2000). "Chiral switches." Lancet **355**(9209): 1085-1087.
Ustun, T. B., J. L. Ayuso-Mateos, et al. (2004). "Global burden of depressive disorders in the year 2000." British Journal of Psychiatry **184**: 386-392.
Vaidya, V. A., J. A. Siuciak, et al. (1999). "Hippocampal mossy fiber sprouting induced by chronic electroconvulsive seizures." Neuroscience **89**(1): 157-166.
van Baardewijk, M., P. M. J. Vis, et al. (2005). "Cost effectiveness of duloxetine compared with venlafaxine-XR in the treatment of major depressive disorder." Current Medical Research and Opinion **21**(8): 1271-1279.
van Loon, J., L. Z. Horvath, et al. (2002). "Cost-effectiveness of mirtazapine compared to fluoxetine in the treatment of moderate and severe depression in Hungary." Value in Health **5**(6): 517-517.
Videbech, P. and B. Ravnkilde (2004). "Hippocampal volume and depression: A meta-analysis of MRI studies." American Journal of Psychiatry **161**(11):

1957-1966.
Wade, A., N. Despiegel, et al. (2002a). "Depression in primary care patients: Improvement during long- term escitalopram treatment." European Neuropsychopharmacology **12**: S232-S233.
Wade, A., O. M. Lemming, et al. (2002b). "Escitalopram 10 mg/day is effective and well tolerated in a placebo-controlled study in depression in primary care." International Clinical Psychopharmacology **17**(3): 95-102.
Wade, A. G., L. Toumi, et al. (2005b). "A pharmacoeconomic evaluation of escitalopram versus citalopram in the treatment of severe depression in the United Kingdom." Clinical Therapeutics **27**(4): 486-496.
Wade, A. G., I. Toumi, et al. (2005a). "A probabilistic cost-effectiveness analysis of escitalopram, generic citalopram and venlafaxine as a first-line treatment of major depressive disorder in the UK." Current Medical Research and Opinion **21**(4): 631-641.
Weihrauch, T. and T. Gauler (1999). "Placebo - efficacy and adverse effects in controlled trials." Arzneimittelforschung **49**(5): 385-393.
Weinstein, M. C., B. O'Brien, et al. (2003). "Principles of good practice for decision analytic modeling in health-care evaluation: Report of the ISPOR task force on good research practices-modeling studies." Value in Health **6**(1): 9-17.
Wells, K. B., A. Stewart, et al. (1989). "The Functioning and Well-Being of Depressed-Patients - Results from the Medical Outcomes Study." Jama-Journal of the American Medical Association **262**(7): 914-919.
Wennogle, L. P. and L. R. Meyerson (1985). "Serotonin Uptake Inhibitors Differentially Modulate High- Affinity Imipramine Dissociation in Human-Platelet Membranes." Life Sciences **36**(16): 1541-1550.
WHO (1992). International Statistical Classification of Diseases and Related Health Problems (ICD-10). Geneva, World Health Organisation.
Wiborg, O. and C. Sanchez (2003). "R-citalopram decreases the association of [^{3}H]-S-citalopram with the human serotonin transporter by an allosteric mechanism." European Neuropsychopharmacology **13**: S217.
Wickberg, B. and C. P. Hwang (1996). "The Edinburgh Postnatal Depression Scale: validation on a Swedish community sample." Acta Psychiatrica Scandinavica **94**(3): 181-4.
Wilde, M. I. and R. Whittington (1995). "Paroxetine - a Pharmacoeconomic Evaluation of Its Use in Depression." Pharmacoeconomics **8**(1): 62-81.
Wilkinson, G., C. Croft-Jeffreys, et al. (1990). "QALYs in psychiatric care?" Psychiatric Bull **14**: 582-585.
Woods, S. W. and J. A. Rizzo (1997). "Cost-effectiveness of antidepressant treatment reassessed." British Journal of Psychiatry **170**: 257-263.
Ziegler, V. E., D. A. Meyer, et al. (1978). "Reliability of Video Taped Hamilton Ratings." Biological Psychiatry **13**(1): 119-122.
Zigmond, A. S. and R. P. Snaith (1983). "The Hospital Anxiety and Depression Scale." Acta Psychiatrica Scandinavica **67**(6): 361-370.
Zimbroff, D. L., D. Li, et al. (2004). "Escitalopram treatment of SSRI non-responders can lead to remission in patients who fail initial SSRI therapy." International Journal of Neuropsychopharmacology **7**: S348-S348.
Zung, W. (1965). "A Self-rating Scale." Arch Gen Psychiatry **12**: 63-70.

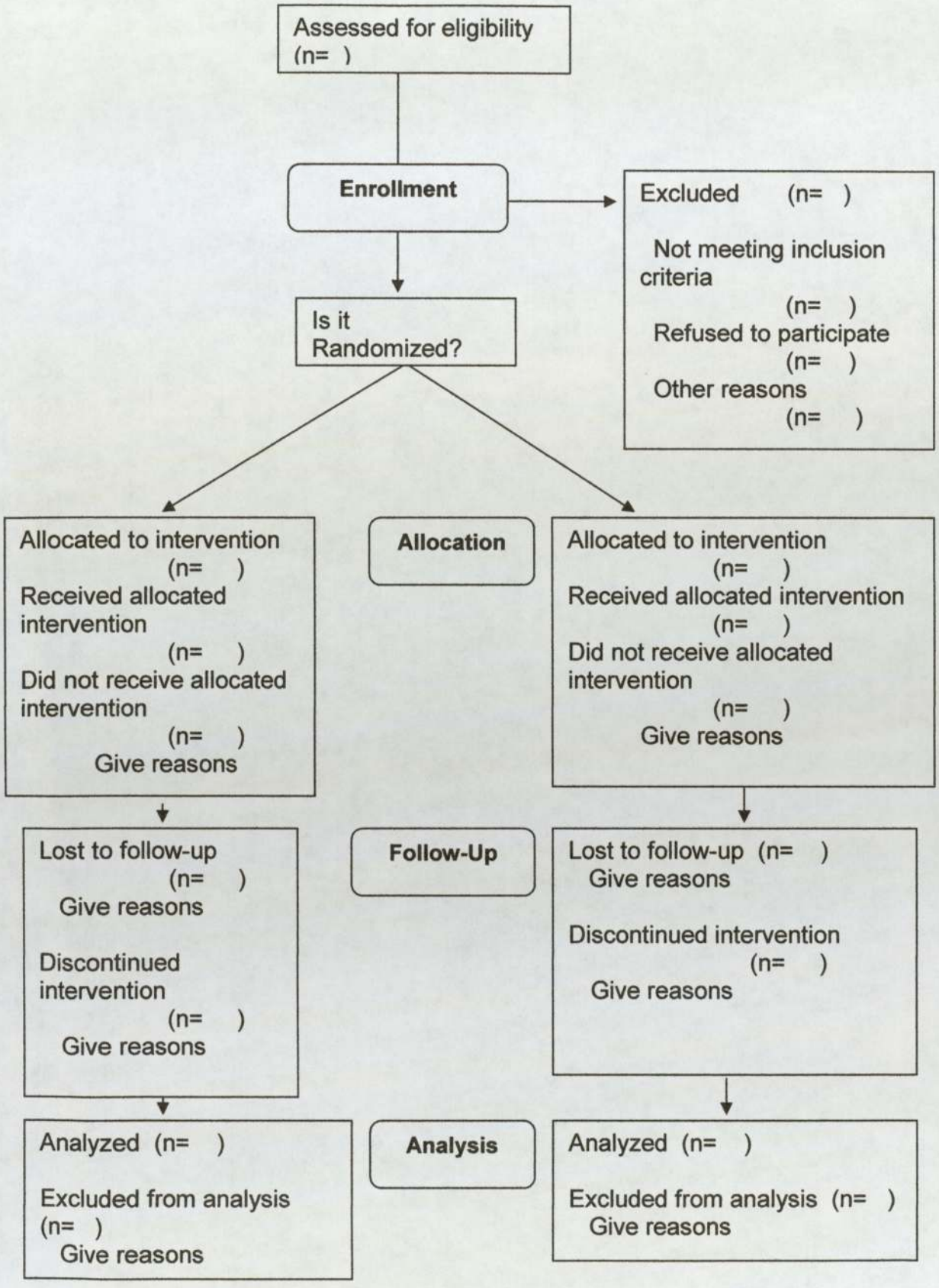
APPENDIX 1

CONSORT **Checklist of items to include when reporting a randomized trial**

PAPER SECTION And topic	Item	Description	Reported on Page #
TITLE & ABSTRACT	1	How participants were allocated to interventions (*e.g.*, "random allocation", "randomized", or "randomly assigned").	
INTRODUCTION Background	2	Scientific background and explanation of rationale.	
METHODS Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.	
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	
Objectives	5	Specific objectives and hypotheses.	
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (*e.g.*, multiple observations, training of assessors).	
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	
Randomization -- Sequence generation	8	Method used to generate the random allocation sequence, including details of any restrictions (*e.g.*, blocking, stratification)	
Randomization -- Allocation concealment	9	Method used to implement the random allocation sequence (*e.g.*, numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.	
Randomization -- Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. When relevant, how the success of blinding was evaluated.	
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); Methods for additional analyses, such as subgroup analyses and adjusted analyses.	
RESULTS Participant flow	13	Flow of participants through each stage (a diagram is strongly recommended). Specifically, for each group report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome. Describe protocol deviations from study as planned, together with reasons.	
Recruitment	14	Dates defining the periods of recruitment and follow-up.	
Baseline data	15	Baseline demographic and clinical characteristics of each group.	
Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat". State the results in absolute numbers when feasible (*e.g.*, 10/20, not 50%).	
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision (*e.g.*, 95% confidence interval).	
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-specified and those exploratory.	

Adverse events	19	All important adverse events or side effects in each intervention group.	
DISCUSSION Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision and the dangers associated with multiplicity of analyses and outcomes.	
Generalizability	21	Generalizability (external validity) of the trial findings.	
Overall evidence	22	General interpretation of the results in the context of current evidence.	

APPENDIX 2 The Consort E-Flowchart



APPENDIX 3 Data from Escitalopram RCTs evaluated by CONSORT Criteria

CONSORT Checklist of items to include when reporting a randomized trial

Trial Ref: Bielski, R. J., D. Ventura, et al. (2004). "A double-blind comparison of escitalopram and venlafaxine extended release in the treatment of major depressive disorder." Journal of Clinical Psychiatry **65**(9): 1190-1196

Topic	Item	Description	Reported on Page #
TITLE/ ABSTRACT	1	How participants were allocated to interventions	1190
Background	2	Scientific background and explanation of rationale.	1190
Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.	1191
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	1191
Objectives	5	Specific objectives and hypotheses.	1191
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements	1192
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	-
Randomization -- Sequence generation	8	Method used to generate the random allocation sequence, including details of any restrictions	-
Randomization -- Allocation concealment	9	Method used to implement the random allocation sequence	-
Randomization -- Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	-
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. When relevant, how the success of blinding was evaluated.	1192
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); Methods for additional analyses,	1192
Participant flow	13	Flow of participants through each stage. Describe protocol deviations from study as planned, together with reasons.	1192
Recruitment	14	Dates defining the periods of recruitment and follow-up.	-
Baseline data	15	Baseline demographic and clinical characteristics of each group.	1192
Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat".	1192
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision	1193 (no ES)
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed,	1194
Adverse events	19	All important adverse events or side effects in each intervention group.	1194
Interpretation	20	Interpretation of the results	1194-95
Generalizability	21	Generalizability (external validity) of the trial findings.	-
Overall evidence	22	General interpretation of the results in the context of current evidence.	-

CONSORT Checklist of items to include when reporting a randomized trial

Trial Ref: Burke, W. J., I. Gergel, et al. (2002). "Fixed-dose trial of the single isomer SSRI escitalopram in depressed outpatients." Journal of Clinical Psychiatry **63**(4): 331-336

Topic	Item	Description	Reported on Page #
TITLE/ ABSTRACT	1	How participants were allocated to interventions	331
Background	2	Scientific background and explanation of rationale.	331
Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.	332
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	332
Objectives	5	Specific objectives and hypotheses.	332
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements	333
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	-
Randomization -- Sequence generation	8	Method used to generate the random allocation sequence, including details of any restrictions	-
Randomization -- Allocation concealment	9	Method used to implement the random allocation sequence	-
Randomization -- Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	-
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. When relevant, how the success of blinding was evaluated.	-
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); Methods for additional analyses,	332
Participant flow	13	Flow of participants through each stage. Describe protocol deviations from study as planned, together with reasons.	333
Recruitment	14	Dates defining the periods of recruitment and follow-up.	-
Baseline data	15	Baseline demographic and clinical characteristics of each group.	333
Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat".	333
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision	333-4 (no ES)
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed,	334-5
Adverse events	19	All important adverse events or side effects in each intervention group.	335
Interpretation	20	Interpretation of the results	335-6
Generalizability	21	Generalizability (external validity) of the trial findings.	336
Overall evidence	22	General interpretation of the results in the context of current evidence.	336

CONSORT Checklist of items to include when reporting a randomized trial

Trial Ref: Lepola, U. M., H. Loft, et al. (2003). "Escitalopram (10-20 mg/day) is effective and well tolerated in a placebo-controlled study in depression in primary care." International Clinical Psychopharmacology **18**(4): 211-217

Topic	Item	Description	Reported on Page #
TITLE/ ABSTRACT	1	How participants were allocated to interventions	211
Background	2	Scientific background and explanation of rationale.	211
Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.	212
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	212
Objectives	5	Specific objectives and hypotheses.	-
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements	212
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	212
Randomization -- Sequence generation	8	Method used to generate the random allocation sequence, including details of any restrictions	-
Randomization -- Allocation concealment	9	Method used to implement the random allocation sequence	-
Randomization -- Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	-
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. When relevant, how the success of blinding was evaluated.	-
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); Methods for additional analyses,	212
Participant flow	13	Flow of participants through each stage. Describe protocol deviations from study as planned, together with reasons.	212-3
Recruitment	14	Dates defining the periods of recruitment and follow-up.	-
Baseline data	15	Baseline demographic and clinical characteristics of each group.	213
Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat".	212
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision	213-4 (no ES)
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed,	213-4
Adverse events	19	All important adverse events or side effects in each intervention group.	214
Interpretation	20	Interpretation of the results	215-6
Generalizability	21	Generalizability (external validity) of the trial findings.	-
Overall evidence	22	General interpretation of the results in the context of current evidence.	-

CONSORT Checklist of items to include when reporting a randomized trial

Trial Ref: Montgomery, S. A., H. Loft, et al. (2001). "Escitalopram (S-Enantiomer of citalopram): Clinical efficacy and onset of action predicted from a rat model." Pharmacology & Toxicology **88**(5): 282-286

Topic	Item	Description	Reported on Page #
TITLE/ ABSTRACT	1	How participants were allocated to interventions	282
Background	2	Scientific background and explanation of rationale.	282
Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.	283
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	283
Objectives	5	Specific objectives and hypotheses.	282
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements	283
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	283
Randomization -- Sequence generation	8	Method used to generate the random allocation sequence, including details of any restrictions	-
Randomization -- Allocation concealment	9	Method used to implement the random allocation sequence	-
Randomization -- Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	-
- Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. When relevant, how the success of blinding was evaluated.	-
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); Methods for additional analyses,	283
Participant flow	13	Flow of participants through each stage. Describe protocol deviations from study as planned, together with reasons.	284
Recruitment	14	Dates defining the periods of recruitment and follow-up.	-
Baseline data	15	Baseline demographic and clinical characteristics of each group.	-
Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat".	284
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision	285
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed,	-
Adverse events	19	All important adverse events or side effects in each intervention group.	285
Interpretation	20	Interpretation of the results	285
Generalizability	21	Generalizability (external validity) of the trial findings.	-
Overall evidence	22	General interpretation of the results in the context of current evidence.	-

CONSORT Checklist of items to include when reporting a randomized trial

Trial Ref: Montgomery, S. A., A. K. T. Huusom, et al. (2004). "A randomised study comparing escitalopram with venlafaxine XR in primary care patients with major depressive disorder." Neuropsychobiology **50**(1): 57-64

Topic	Item	Description	Reported on Page #
TITLE/ ABSTRACT	1	How participants were allocated to interventions	57
Background	2	Scientific background and explanation of rationale.	57
Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.	58
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	58
Objectives	5	Specific objectives and hypotheses.	58 (aim)
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements	58 (no methods)
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	58 (only sample size)
Randomization -- Sequence generation	8	Method used to generate the random allocation sequence, including details of any restrictions	-
Randomization -- Allocation concealment	9	Method used to implement the random allocation sequence	-
Randomization -- Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	-
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. When relevant, how the success of blinding was evaluated.	58
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); Methods for additional analyses,	58
Participant flow	13	Flow of participants through each stage. Describe protocol deviations from study as planned, together with reasons.	59-60
Recruitment	14	Dates defining the periods of recruitment and follow-up.	-
Baseline data	15	Baseline demographic and clinical characteristics of each group.	59
Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat".	59
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision	60
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed,	59-60
Adverse events	19	All important adverse events or side effects in each intervention group.	60
Interpretation	20	Interpretation of the results	62
Generalizability	21	Generalizability (external validity) of the trial findings.	-
Overall evidence	22	General interpretation of the results in the context of current evidence.	63

CONSORT Checklist of items to include when reporting a randomized trial

Trial Ref: Moore, N., H. Verdoux, et al. (2005). "Prospective, multicentre, randomized, double-blind study of the efficacy of escitalopram versus citalopram in outpatient treatment of major depressive disorder." International Clinical Psychopharmacology **20**(3): 131-137

Topic	Item	Description	Reported on Page #
TITLE/ ABSTRACT	1	How participants were allocated to interventions	131
Background	2	Scientific background and explanation of rationale.	131
Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.	132
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	132
Objectives	5	Specific objectives and hypotheses.	131
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements	132 (training used)
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	132
Randomization -- Sequence generation	8	Method used to generate the random allocation sequence, including details of any restrictions	132
Randomization -- Allocation concealment	9	Method used to implement the random allocation sequence	-
Randomization -- Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	-
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. When relevant, how the success of blinding was evaluated.	132
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); Methods for additional analyses,	132-3
Participant flow	13	Flow of participants through each stage. Describe protocol deviations from study as planned, together with reasons.	133 (flow diagram)
Recruitment	14	Dates defining the periods of recruitment and follow-up.	-
Baseline data	15	Baseline demographic and clinical characteristics of each group.	134
Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat".	134
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision	134 (no ES)
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed,	134
Adverse events	19	All important adverse events or side effects in each intervention group.	135
Interpretation	20	Interpretation of the results	135-6
Generalizability	21	Generalizability (external validity) of the trial findings.	137
Overall evidence	22	General interpretation of the results in the context of current evidence.	136-7

CONSORT Checklist of items to include when reporting a randomized trial

Trial Ref: Wade, A., O. M. Lemming, et al. (2002b). "Escitalopram 10 mg/day is effective and well tolerated in a placebo-controlled study in depression in primary care." International Clinical Psychopharmacology **17**(3): 95-102

Topic	Item	Description	Reported on Page #
TITLE/ ABSTRACT	1	How participants were allocated to interventions	95
Background	2	Scientific background and explanation of rationale.	95-6
Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.	96
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	96
Objectives	5	Specific objectives and hypotheses.	-
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements	96
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	-
Randomization -- Sequence generation	8	Method used to generate the random allocation sequence, including details of any restrictions	-
Randomization -- Allocation concealment	9	Method used to implement the random allocation sequence	-
Randomization -- Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	-
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. When relevant, how the success of blinding was evaluated.	-
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); Methods for additional analyses,	96
Participant flow	13	Flow of participants through each stage. Describe protocol deviations from study as planned, together with reasons.	97, 98
Recruitment	14	Dates defining the periods of recruitment and follow-up.	-
Baseline data	15	Baseline demographic and clinical characteristics of each group.	97
Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat".	97
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision	97 (no ES)
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed,	99
Adverse events	19	All important adverse events or side effects in each intervention group.	99-100
Interpretation	20	Interpretation of the results	101
Generalizability	21	Generalizability (external validity) of the trial findings.	101
Overall evidence	22	General interpretation of the results in the context of current evidence.	101

APPENDIX 4 Contingency Tables for Escitalopram

2x2 Contingency Table for Escitalopram Trials - Responders (ITT; Active Comparator)

Stratum	Author	Year	Comparator	a	b	c	d	a+c (ESC)	b+d (Comp)	OR	RR	RD	NNT	LN(OR)	Var(LN(OR))	SE(LN(OR))	w
1	Alexopoulos	2004	Sertraline	78	70	26	37	104	107	1.5857	1.146	-0.1	10	0.4610	0.0926	0.3043	10.80
2	Bielski	2004	Venlafaxine	57	47	40	51	97	98	1.5463	1.225	-0.11	9	0.4358	0.0834	0.2888	11.99
3	Burke (10mg)	2002	Citalopram	59	57	59	68	118	125	1.193	1.096	-0.04	23	0.1765	0.0661	0.2572	15.12
4	Burke (20mg)	2002	Citalopram	63	57	60	68	123	125	1.2526	1.123	-0.06	18	0.2252	0.0648	0.2545	15.43
5	Colonna	2002	Citalopram	70	63	15	22	85	85	1.6296	1.111	-0.08	12	0.4884	0.1423	0.3772	7.03
6	Lepola	2003	Citalopram	99	84	56	75	155	159	1.5784	1.209	-0.11	9	0.4564	0.0532	0.2306	18.80
7	Montgomery	2004	Venlafaxine	113	113	33	29	146	142	0.8788	0.973	0.022	-46	-0.1292	0.0825	0.2872	12.12
8	Moore	2005	Citalopram	105	87	33	55	138	142	2.0115	1.242	-0.15	7	0.6989	0.0695	0.2636	14.39

Stratum	Lower CI (in terms of OR)	Upper CI	M-H Wt	Fixed Effect (Yp)	Fixed Effect (pooled) Lower CI	Fixed Effect (pooled) Upper CI	Variance Yp (Vp)	Q	D-L (V_I)	I^2	Random Effect Point (y^*_p)	Random Effect Upper CI	Random Effect Lower CI
1	0.87	2.88	8.625592										
2	0.88	2.72	9.641026										
3	0.72	1.97	13.83951										
4	0.76	2.06	13.79032										
5	0.78	3.41	5.558824										
6	1.00	2.48	14.98089		0.158034	0.539360					0.348495	0.470343	-0.615774
7	0.50	1.54	12.94792		1.171206	1.714909					as OR =	1.600544	0.540223
8	1.20	3.37	10.25357	0.3486969	point est =	1.417220	0.009463	5.80	0.08	21		point est =	1.416934

APPENDIX 4 Contingency Tables for Escitalopram (cont'd)

2x2 Contingency Table for Escitalopram Trials - Remitters (ITT; Active Comparator)

Stratum	Author	Year	Comparator	a	b	c	d	a+c (ESC)	b+d (Comp)	OR	RR	RD	NNT	LN(OR)	Var(LN(OR))	SE(LN(OR))	w (in
1	Alexopoulos	2004	Sertraline	63	64	41	44	104	108	1.0564	1.022	-0.01	76	0.054869	0.078616	0.280385	12.72
2	Bielski	2004	Venlafaxine	40	36	57	62	97	98	1.2086	1.123	-0.05	22	0.189444	0.086451	0.294025	11.57
3	Colonna	2002	Citalopram	70	60	15	25	85	85	1.9444	1.167	-0.12	9	0.664976	0.137619	0.370970	7.27
4	Lepola	2003	Citalopram	81	68	74	91	155	159	1.4648	1.222	-0.09	11	0.381736	0.051554	0.227055	19.40
5	Montgomery	2004	Venlafaxine	102	99	44	43	146	142	1.0069	1.002	-0	691	0.006863	0.065888	0.256687	15.18
6	Moore	2005	Citalopram	75	61	63	81	138	142	1.5808	1.265	-0.11	9	0.457929	0.057945	0.240719	17.26

Stratum	Lower CI (in terms of OR)	Upper CI	M-H Wt	Fixed Effect (Yp)	Fixed Effect (pooled) Lower CI	Fixed Effect (pooled) Upper CI	Variance Yp (Vp)	Q	D-L (V_I)	I^2	Random Effect Point (y^*_p)	Random Effect Upper CI	Random Effect Lower CI
1	0.61	1.83	12.37736										
2	0.68	2.15	10.52308										
3	0.94	4.02	5.294118										
4	0.94	2.29	16.02548		0.06	0.49					0.279612	0.403417	-0.875459
5	0.61	1.67	15.125		1.0647916	1.6356796					as OR =	1.496932	0.416671
6	0.99	2.53	13.725	0.2774187	point est =	1.3197189	0.0120	3.70	0.07	35		point est =	1.322616