Systematic Review of Caregiver Support
for the Treatment of A Postpartum Depression

by

Karen L. Ray

A thesis submitted in conformity with the requirements
for the Degree of Master of Science
Graduate Department of Nursing
in the University of Toronto

© Copyright by Karen L. Ray, 1999
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

0-612-45551-3

Canada
A Systematic Review of Caregiver Support for the Treatment of Postpartum Depression

A thesis submitted in conformity with the requirements for the degree of Master of Science Graduate Department of Nursing in the University of Toronto, 1999

By
Karen L. Ray

Abstract

The Cochrane Collaboration aims to prepare and maintain systematic reviews of randomized controlled trials of the effects of health care, and to make this information readily available to decision-makers at all levels of health care systems. The objective of this systematic review for the Cochrane Library was to obtain unbiased estimates of the effectiveness of professional and/or social interventions in the treatment of postpartum depression on indicators of maternal, infant and family morbidity. Relevant trials were identified in the Pregnancy and Childbirth Group's Specialized Register of Controlled Trials and through Index Medicus. Studies were included if the purpose was clearly defined as evaluation of support as a treatment for postpartum depression, with acceptable methods of randomization. Results suggest that treatment of postpartum depression with support is beneficial, but these results must be viewed with caution due to the small sample sizes and wide confidence intervals. As only two trials have evaluated the use of social support in the treatment of postpartum depression, further study in this area is needed.
ACKNOWLEDGEMENTS

I wish to thank my committee chairperson, Dr. Ellen Hodnett, for her patience, guidance and sense of humour during the development and preparation of this thesis.
I would also like to thank the other members of my committee, Professor Ruth Gallop and Dr. Beverley Chalmers, for their support and constructive criticism.

To the Editors of the Pregnancy and Childbirth Group of the Cochrane Collaboration for allowing me to conduct this review in perinatal health and for supporting me in this endeavor.

To the authors of the trials in this review for providing me with additional information not found in the published reports. I would also like to thank the women who took part in the trials used in this review. Without their willingness to help others understand their condition, this study would not have been possible.

Lastly, I wish to thank my family and friends for their unrelenting support. Especially to my children, Matthew and Melissa, thank you for your understanding and assistance in making my dream a reality.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Abstract</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>Table of Appendices</td>
<td>vi</td>
</tr>
<tr>
<td>Chapter I THE PROBLEM</td>
<td></td>
</tr>
<tr>
<td>Literature Review</td>
<td>3</td>
</tr>
<tr>
<td>Postpartum Depression: Definition and Descriptive Studies</td>
<td>3</td>
</tr>
<tr>
<td>Measurement Issues in Postpartum Depression</td>
<td>6</td>
</tr>
<tr>
<td>Etiology of Postpartum Depression</td>
<td>8</td>
</tr>
<tr>
<td>Traditional Explanations:</td>
<td>8</td>
</tr>
<tr>
<td>Medical Model</td>
<td>8</td>
</tr>
<tr>
<td>Personality and Psychoanalytic Theories</td>
<td>9</td>
</tr>
<tr>
<td>Social Models</td>
<td>10</td>
</tr>
<tr>
<td>Treatments for Postpartum Depression</td>
<td>11</td>
</tr>
<tr>
<td>Medication</td>
<td>11</td>
</tr>
<tr>
<td>Counselling:</td>
<td>12</td>
</tr>
<tr>
<td>Support Groups</td>
<td>12</td>
</tr>
<tr>
<td>Group Psychotherapy</td>
<td>12</td>
</tr>
<tr>
<td>Individual Psychotherapy</td>
<td>13</td>
</tr>
<tr>
<td>Theoretical Basis for Social Support as a Treatment Modality</td>
<td>13</td>
</tr>
<tr>
<td>Social Support in the Perinatal Period</td>
<td>17</td>
</tr>
<tr>
<td>Summary</td>
<td>18</td>
</tr>
</tbody>
</table>

iv
Chapter II

THE COCHRANE REVIEW

Abstract

Background

Types of Participants

Types of Studies

Types of Interventions

Types of Outcome Measures

Search Strategies

Methods of the Review

Assessment of Methodological Quality

Extraction of the Data

Data Synthesis

Methodological Quality of Included Studies

Results

Discussion

Implications for Practice

Implications for Research

References of Included Trials

Additional References

Characteristics of Included Trials

Table of Comparisons
CHAPTER III CONCLUSIONS

References 31

Appendices 35

Appendix A Overview of the Cochrane Collaboration 48
Appendix B Search Strategies 58
Appendix C Critical Appraisal of Studies 63
CHAPTER I THE PROBLEM

Postpartum depression has become a broad, illness-dominated definition for unhappiness following childbirth (Romito, 1989). Complex factors have contributed to varied reported rates and a lack of consensus on definition, symptoms, and causes of postpartum depression. Theoretical and measurement issues further complicate its investigation and may contribute to the over or under identification of this problem.

Estimates of the incidence of postpartum depression range from seven to thirty percent, depending on the criteria and instruments used in its assessment. In addition, duration has been poorly documented due to a lack of longitudinal studies (Romito, 1989). Women experiencing postpartum depression appear to suffer disabling symptoms including excessive fatigue, insomnia, inability to cope, and suicidal ideation (Affonso & Domino, 1984; Pitt, 1968). This morbidity may have a profound effect on the mother, infant, and family (Cogill et al., 1986; Murray, 1992; Oates, 1994).

Controversy exists as to the causes of postpartum depression. The medical model proposes that depression results from changes in hormone levels during the postpartum period, while personality and psychoanalytic theories suggest that pre-existing factors in the woman's personality and history make her vulnerable to postpartum depression. These theories remain unproven and ignore social factors, despite little evidence that postpartum depression exists in non-industrialized societies (Romito, 1989).

Several treatment approaches have been suggested for postpartum depression, including the use of social support. The concept of social support is rooted in the belief that individuals need a meaningful relationship with others. Crucial aspects of this relationship are
the perception or the experience of feeling that one matters to others (Stemp, Turner & Noh, 1986). Social support appears to be beneficial during the perinatal period. Supportive relationships during the perinatal period may enhance feelings of well-being, personal control, and positive affect, and thereby help women to perceive pregnancy-related changes as less stressful (Norbeck & Anderson, 1989). Support to women during labour has been shown to reduce length of labour, operative delivery, and other medical interventions (Hodnett, 1997). During the postpartum period, home-based social support for socially disadvantaged mothers and babies has been associated with reduced hospital admissions, less use of the hospital emergency department, and fewer children with incomplete immunizations. Mothers' reports of unhappiness were also reduced (Hodnett & Roberts, 1997).

No systematic review was found concerning the effectiveness of support as a treatment for postpartum depression. Cochrane recognized that people who want to make informed decisions in health care did not have ready access to reliable evidence. He advocated for a method of obtaining a critical summary of evidence that periodically reviewed all relevant randomized controlled trials by specialty. The Cochrane Collaboration was developed to support the need for systematic, up-to-date reviews and to make information available to decision makers at all levels of the health care system (Cochrane Library, 1996; Appendix A).

A Cochrane Review begins when a topic is identified that requires a summary of evidence. A protocol is developed to outline the parameters of the trials accepted into the review. Relevant trials are obtained through search strategies (Appendix B) and evaluated using a standardized method of analysis (Appendix C). To ensure consistency among reviewers, methods for reviewing trials are provided in the Cochrane Handbook and via
consultation with editorial teams (Cochrane Library, 1996).

It should be noted that, although the Cochrane Collaboration provides a method for reviewing trials and making more informed decisions, the evidence is based on randomized controlled trials only. The randomized controlled trial study design attempts to decrease bias to provide the most reliable evidence possible, but in so doing may limit the type of information available to the caregiver.

In keeping with the methods of the Cochrane Collaboration, this thesis reviews literature regarding postpartum depression and social support (Chapter I) and presents a systematic review of the randomized controlled trials that evaluated social support as a treatment for postpartum depression (Chapter II).

Literature Review

Postpartum Depression: Definition and Descriptive Studies

Postpartum depression was first described by Pitt (1968) as an "atypical" depression that followed childbirth. Since then, the study of postpartum depression has been complicated by varied reported rates and conflicting views on definition, symptoms, and its effect on families. The lack of measurement tools specific to postpartum depression and a limited understanding of the etiology have further hampered its investigation.

Postpartum psychiatric syndromes are generally described in three categories: "postpartum blues", postpartum depression, and postpartum psychosis. "Postpartum blues", experienced by 50 to 80 percent of postpartum women, is characterized by transitory emotional lability within two to four days postpartum (Affonso and Domino, 1984; Pitt, 1964;
Romito, 1989). At the other end of the spectrum is postpartum psychosis, a relatively rare but severe disorder, with an incidence of 0.1-0.2 percent. Postpartum psychosis, which may manifest as manic-depressive illness, primary depression, or schizophrenia, occurs in the first three months postpartum and requires intensive medical therapy (Affonso & Domino, 1984; Kendell, 1985; Pitt, 1968; Romito, 1989).

What lies between these two extremes is described as postpartum depression (Pitt, 1968). Postpartum depression is classified by the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 1994), as similar to minor or major depressive episodes in the general population, with onset four weeks after delivery of a child. Postpartum depression has an estimated incidence of 3 to 27 percent, although it may often go undetected by the medical profession (Holden, 1994). The duration of postpartum depression is questionable; few longitudinal studies have been done and postpartum depression appears to be present for as long as it is followed (Cogill, Caplan, Alexandra, Robson, & Kumar, 1986; Cutrona, 1983; Kumar and Robson, 1984).

Women experiencing postpartum depression suffer disabling symptoms. Symptoms include disinterest in usual activities, excessive fatigue, insomnia, appetite disturbances, and suicidal ideation (Affonso & Arizmendi, 1986; Pitt, 1968). Women themselves describe the experience with similar themes such as: loneliness, contemplating death to end their nightmare, obsessive thoughts of being a bad mother, grief over loss of self, life seems empty of all previous interests and goals, fear and guilt over pondering harm to infants, inability to concentrate, uncontrolled anxiety and loss of control, and insecurity (Beck, 1992; Nicholson, 1990).
Postpartum depression may have a profound effect on mother, infant, and family. Mothers with postpartum depression may experience little enjoyment of their children during the postpartum period and long-term relationships may be affected (Oates, 1994). Depressed woman may be unresponsive to infant cues (Beck, 1995), which, in turn, may jeopardize infant development. Children raised by women who have suffered postpartum depression may have lower cognitive performance (Cogill et al., 1986; Hay & Kumar, 1995) and behavioral problems (Murray, 1992). Plans to have more children (Peindl, Wisner, Zolnik, & Hanusa, 1995) and the marital relationship may also be disrupted (Holden, 1991; 1994).

Therefore, postpartum depression is a serious condition that results in significant social, emotional, and psychological morbidity for women and their families. Yet practitioners are unsure of the interventions that would be helpful to treat postpartum depression. Much of this uncertainty hinges on the fact that the etiology of postpartum depression remains unclear. In fact, although many authors describe postpartum depression, others question its existence (Deacon, 1983; Romito, 1989). Some authors believe that depression occurring in the postpartum period may actually be on the continuum of unhappiness that plagues women generally in our society (Chalmers & Chalmers, 1986; O'Hara, Lewis, Schlechte, & Varner, 1991; Romito, 1989; Weissman & Olfson, 1995).

To assist in clarifying these issues, Romito (1989) conducted a systematic review of the literature on the definition and etiology of postpartum depression. She argued that there is little benefit to labelling depression observed during the postpartum period as a separate entity from depression occurring at other times, for similar rates of depression are found in mothers at various intervals throughout childrearing (Richman, 1974; Moss & Plewis, 1977).
Romito (1989) also concluded that postpartum depression may be defined and artificially constructed by the instruments used to measure it, or falsely viewed against the "normal". Therefore, it should not be surprising that many factors have been associated with postpartum depression when different authors study different samples of women, using different instruments and criteria for depression, at different times after birth.

**Measurement Issues in Postpartum Depression**

Since Pitt (1968) first described postpartum depression, authors have attempted to measure postpartum depression as a discrete entity. Measurement has been problematic. Studies with methodological limitations have reduced the ability to identify differences within and between groups and generalize research findings (Burns & Groves, 1993). Research designs and sampling methods have allowed for selection, information, confounding, reverse causality, and recall biases. In many studies, sample sizes were small (for example, Gennaro, 1988; Laizner & Jeans, 1990; Mike, McGovern, Kochevar & Roberts, 1994; O'Hara, 1986), and subjects were primarily white, educated, middle-class women (for example, Gjerdingen & Chaloner, 1994; O'Hara, 1986). The majority of studies used descriptive correlational designs (for example, Affonso & Arizmendi, 1986; Griepsma, Marcollo, Casey, Cherry, Vary & Walton, 1994), and often ignored risk factors such as previous depressive episodes (for example, Beck, Reynolds & Rutowski, 1992; Laizner & Jeans, 1990) and perinatal loss (Laizner & Jeans, 1990; Ugarriza, 1995).

The lack of a standardized method for measuring the incidence, severity, or duration of postpartum depression has further hindered research. Numerous instruments have been
used to identify postpartum depression including the Beck Depression Inventory (BDI) (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), the Hamilton Depression Scale (HDS) (Hamilton, 1976) and the Center for Epidemiologic Studies Depression Scale (CES-D) (Radloff, 1977). These instruments have been shown to be reliable and valid when used to identify depression in the general population, but symptoms of the normal postpartum process, such as sleep disturbances and loss of interest in usual activities are also indices of depressive reactions (Logsdon, McBride, Birkimer, 1994). In addition, women completing these instruments may be concerned about admitting they are depressed, for such disclosures may cast doubts about their capabilities as successful mothers (Elliott, 1994).

One of the more recent advancements in the measurement of postpartum depression is the development of the Edinburgh Postnatal Depression Scale (EPDS) (Cox, Holden & Sagovsky, 1987). This screening tool was designed to be used in primary care settings by personnel without specialized psychiatric training. The ten-item questionnaire can be completed in minutes and appears to be acceptable to women as indicated by good response rates (for example, Griepehna et al., 1994 (62%); Murray & Carothers, 1990 (95%); Stamp and Crowther, 1994 (95% at 6 weeks and 82% at 6 months postpartum). Reliability and validity studies have found good evidence of sensitivity and specificity (Cox, Holden & Sagovsky, 1987 (86%, 78%); Murray & Carothers, 1990 (67.7%, 95.7%); Harris, Huckle, Thomas, Johns, & Fung, 1989 (95%, 93%), with differences in scores likely to be due to sampling variation between studies.
Etiology of Postpartum Depression

Once identified as depressed, women have been questioned about factors that may have been associated with the onset of depression. Many variables, including past psychiatric history, obstetrical difficulties and complications, age, anxiety during pregnancy, parity, and various social situations have been noted (Affonso & Domino, 1984; Chalmers & Chalmers, 1986; Kendell, 1985; Kumar & Robson, 1984; Romito, 1989). One variable found to have a consistent association with postpartum depression is marital conflict. Marital conflict is described as disruption in the marital relationship, or the perception of inadequate support from a spouse (Arizmendi & Affonso, 1984), and was found to be a factor associated with postpartum depression in every one of eleven studies that included it as a variable. It should be noted that this finding may represent a circular explanation for postpartum depression. Although it is possible that postpartum depression is caused because of a poor marital relationship, the marital relationship may also be disrupted because of postpartum depression.

In the past, several models were proposed to explain the etiology of postpartum depression. It should be noted that today depression is viewed as having multi-dimensional causes. For purposes of clarity these models are described separately below.

Traditional Explanations: Medical Model

Some authors believe that postpartum depression has a physiological origin, and have described postpartum depression as a disease or illness. To date, no evidence of a direct or indirect link between hormones and postpartum depression has been found (Kendall, 1985). The biophysical theory attributes postpartum depression to hormonal changes or genetic
predisposition, although this theory fails to account for women who do not develop postpartum depression (Affonso & Domino, 1984). The lack of evidence between postpartum depression and a biological link is further supported by other research findings. For example, new fathers have been found to be depressed (Atkinson & Rickel, 1984), and postpartum depression does not appear to be present in all cultural groups (Stern & Kruckman, 1983). Other aspects of the biological domain have not been well researched. The physical consequences that pregnancy, childbirth, lactation, and obstetrical interventions have on the physical and psychological well-being of mothers are ignored by most researchers (Romito, 1989).

**Personality and Psychoanalytic Theories**

As with the medical model, personality and psychoanalytic theories are a reductionist approach to explain postpartum depression. Personality theory proposes that women who are anxious, over-controlled, or have low self-esteem have a greater chance of developing postpartum depression (Romito, 1989). Some personality theories attempt to correlate other affective disorders, such as premenstrual syndrome, with postpartum depression. Psychoanalytic theories hypothesize that postpartum depression results from the emotional regression of the mother due to unresolved childhood conflicts and dependency issues. Both of these theories have been criticized because they cannot be proven empirically (Affonso & Domino, 1984; Romito, 1989).
Social Models

Both the medical and personality/psychoanalytical models exclude the social circumstances in which women must function (Romito, 1989). Often, women must raise children in conditions of economic strife and emotional and physical abuse. In addition, the work involved in being a new parent is substantial and these activities are often boring and isolating (Affonso & Arizmendi, 1986; Romito, 1989). Today, the weight of research findings and the prevailing current opinion is that psychosocial factors are at least as important, if not more important, than biological factors in the causation of most major and minor depressive illnesses in the puerperium (Oates, 1994).

Several studies have suggested a relationship between postpartum emotional problems and a lack of practical and emotional support after childbirth. Although the research had serious methodological limitations such as convenience sampling methods and the lack of randomization, Gordon & Gordon (1959; 1960) were the first researchers to hypothesize such a relationship. Paykel, Emms, Fletcher and Rassaby (1980) also studied the effects of social support on postpartum depression. Using a model in which clinical depression is considered to be an understandable response to adversity, Paykel et al. (1980) found that the presence of vulnerability factors, such as lack of a supportive relationship with a spouse, lowered a woman's self-esteem and increased the likelihood of development of depression. When coupled with a negative life event such as the loss of employment or a loved one, depression may result. Numerous descriptive studies on social support in the postpartum period support this theory and found women suffering postpartum depression rated their spouses as very deficient in both emotional and tangible support (Gjerdingen & Chaloner, 1994; Logsdon,
McBride & Birkimer, 1994; O'Hara, 1986), especially when this lack of help coincided with some negative life event.

Treatments for Postpartum Depression

Although social support has been hypothesized as a treatment modality for postpartum depression, other options have also been proposed. A review of the literature found that treatment has generally focused on the use of medication and counselling.

Medication

Estrogen has been used in the treatment of severe postpartum depression, on the premise that postpartum estrogen withdrawal provokes affective psychosis through interaction with central dopaminergic systems in women with a history of bipolar or schizoaffective psychosis (Wieck, Kumar, Hirst, Marks, Campbell & Checkley, 1991). Results showed an improvement in EPDS scores for the depressed women receiving estrogen supplements, although results should be viewed with caution due to lack of randomization and small sample size (Henderson, Gregoire, Kumar & Studd, 1991).

Other authors have proposed that early treatment with antidepressants prescribed in appropriate dosage may also limit the disabling effects of postpartum depression (Buist, 1993; Cox, 1989; Holden et al., 1989; Millis & Kornblith, 1992). However, compliance with the use of anti-depressants is often poor due to side-effects and the desire to breastfeed (Holden, 1994).
Counselling

Support Groups.

Several types of counselling have been suggested to treat postpartum depression, ranging from self-help groups to psychotherapy (Elliott, 1989). The use of support groups for treatment of postpartum depression has been described by numerous authors (for example, Handford, 1985; Olson, Cutler, & Legault, 1991; Sanderson & Curry, 1996). Perhaps the most widely known group is the Pacific Postpartum Support Society. The group was founded on the premise that, in general, psychiatrists are not knowledgeable about what it means to be at home and to take care of small children. Treatment consists of talking to other women who have experienced the same difficulties, initially by telephone and then through group meetings (Handford, 1985). No evaluation of the program has been reported.

Only one quasi-experimental study was found that attempted to evaluate a postpartum support group. Fleming, Klein, & Corter (1992) compared depressed mothers who attended an eight-week social support group with a group of non-depressed mothers. In their small sample of 44, the researchers found that the postpartum support group did not appear to alleviate depression, and appeared to be detrimental to mothers' self-confidence.

Group Psychotherapy.

Psychotherapy has also been suggested as a treatment for postpartum depression. Gruen (1993) described the use of a psychotherapy group to promote understanding of postpartum depression, provide stress reduction techniques to decrease anxiety, enhance self-esteem, and improve communication skills between couples. Morris (1987) evaluated group
psychotherapy for seven women who had been suffering postpartum depression for more than a year. Weekly meetings for eleven months used psychodrama techniques to facilitate discussion on topics such as marital and sexual problems, and hostility towards the women's parents. A decrease in depression scores was noted. However, the absence of a control group renders it impossible to draw causal inferences.

**Individual Psychotherapy.**

Stuart and O'Hara (1995) used interpersonal psychotherapy (Weissman & Markowitz, 1994) to treat six women with postpartum depression. Therapy focused on the symptoms and current interpersonal problems and continued for twelve weeks. The researchers observed a decline in depression scores, although the absence of a comparison group makes it impossible to draw causal inferences.

**Theoretical Basis for Social Support as a Treatment Modality**

The need for meaningful relationship with others is the assumption underlying the concept of social support. Crucial aspects of this relationship are the perception or the experience of being supported (Stemp, Turner & Noh, 1986). An inverse relationship between supportive interpersonal relationships and psychological morbidity, suicide, and clinical depression have been identified by several authors (Broadhead et al., 1983; Cohen & Wills, 1985; Kessler & McLeod, 1985). This relationship may be particularly important to maintain health during the transition to parenthood. Not only is the transition to parenthood a stressful experience for most parents (Eckenrode & Gore, 1981), but it may cause disruptions in
existing social supports (Gottlieb, 1981).

The lack of a theoretical framework for social support has resulted in varied definitions and methods of measurement (Gottlieb, 1981), although most authors now agree that a definition of social support should include both structural and functional factors (Cohen & Syme, 1985; Cohen & Wills, 1985). Structural measures indicate the range and interconnectedness of extant social support resources (Ganster & Victor, 1988). In contrast, measures of functional social support identify the qualities of social relationships as perceived by the recipient (Cohen & Wills, 1985). Much of the original research on social support focused on structural support measures that described the existence of and interconnections between social ties, such as number of relationships. However, functional supportive measures may have more of an influence on a multidimensional phenomenon such as postpartum depression.

Functional measures describe three types of support: information, tangible assistance, and emotional support. Informational support is the process through which other persons may provide information, advice, and guidance. Information may enhance perceptions of control by providing patients with ways of managing their condition. Tangible support consists of helping in the home, providing transportation, material goods, or lending money. Emotional support involves the verbal and nonverbal communication of caring and concern. Emotional support can help to restore self-esteem or reduce feelings of personal inadequacy by communicating that one is valued and loved. This may permit the expression of feelings that may reduce distress and can lead to greater attention to and improvement in interpersonal relationships (Anderson & Deshaies, 1996; Helgeson & Cohen, 1996; Wills, 1985).
The benefits of functional support may be obtained by the woman with postpartum depression through direct or indirect routes. The main effects hypothesis suggests that social support may directly affect mental and physical health and will thus provide benefit regardless of stress levels. The main effects of support have been primarily found when support is experienced as social integration within a social network that is perceived as helpful (Cohen & Wills, 1985). Alternatively, the buffering hypothesis contends that social support may operate indirectly by buffering an individual's reaction to stress. This model focuses on the perceived availability of social support and is valid only when there is a match between the needs elicited by the stressful event and the functions of support that are perceived to be available (Cohen & Wills, 1985).

Three theories exist for the methods by which direct and indirect support work: physiological, psychological, and behavioural. The physiological action of support may occur in several ways. First, some suggest that humans may be genetically programmed to form social relationships, perhaps to enhance survival (House, Umberson, & Landis, 1988). Second, stress may affect health through two physiological routes: the fight or flight response or suppression of an immune system response (Ganster & Victor, 1988). Bovard (1962) suggested that support may compete with the fight or flight reaction by activating a positive brain system response through the hypothalamus to reverse the hypertensive effects of stress. The precise mechanism appears to be the release of certain neuropeptides, especially beta endorphins, in response to socially supportive stimuli (Ganster & Victor, 1988).

Psychological mechanisms may also be the route through which women with postpartum depression benefit from support. According to this theory, social support results
in changes to the perception of the world in general and the cognitive appraisal of potentially stressful events in particular. Wills (1985) suggested that stress can lower self-esteem, perceptions of self-efficacy, and a sense of social control. Lowered psychological responses can result in depression. The knowledge that support is available is thought to lead to more positive affect, a better psychological state, and improved health (Cohen & Syme, 1985; Kessler & McLeod, 1985).

Social support may also promote health by promoting behavioral changes. Persons with positive support systems are encouraged to engage in healthful behavior, receive useful information that may increase these behaviors, or have such behavior directly facilitated by others, resulting in improved health and well-being (Berkman, 1985; Cohen, 1988; Ganster & Victor, 1988).

Although the type of support and its mechanism are important factors in the understanding of social support, the context in which support is perceived, mobilized, given and taken should also be considered in its measurement (Cohen & Syme, 1985). Intrinsic characteristics of the individual may determine the need or desire for support, while individual personality factors may have a profound influence on how support is received (Ganster & Victor, 1988). Variables such as gender (Levy, 1983; Shumaker & Hill, 1991), culture (Norbeck & Tilden, 1988; Stewart & Tilden, 1995; Zink, 1996), and recipient satisfaction (Hobfoll, Nadler & Leiberman, 1986) may also be important factors to consider when measuring social support.

Some authors also suggest the need to examine the characteristics of the support being provided. For example, who provides the support (Dakof & Taylor, 1990) and the timing of
the support in relation to the onset of an event may influence outcomes (Collins, Dunkel-Schetter, Lobel, & Scrimshaw, 1993). Cohen and Syme (1985) pointed out that short-term supportive interventions may have little impact on outcomes that have long developmental periods.

Finally, the demands required by each particular stressor should be examined individually (Cohen & Wills, 1985). Different stressors, such as caring for a new baby, may vary in the extent to which they benefit from a specific type or aspects of support (Hobfoll et al., 1986). As a result, support may be better detected in studies of specific life events and their related stressors (Kessler & McLeod, 1985).

Social Support in the Perinatal Period

Supportive relationships during the perinatal period may enhance feelings of well-being and personal control, thereby helping women to perceive pregnancy-related changes as less stressful (Norbeck & Anderson, 1989). In particular, low emotional support accompanied with a stressful life event has been related to emotional disequilibrium in the postpartum period (Collins et al., 1993; Norbeck & Tilden, 1983). Examining social support in the context of pregnancy is relatively straightforward because pregnancy is time-limited and has specific outcomes (Collins et al., 1993). Both main effects (Turner, Grindstaff, & Phillips, 1990) and stress-buffering effects of social support (Collins et al., 1993; Norbeck & Tilden, 1983; Nuckolls, Cassel, & Kaplan, 1972) have been demonstrated in studies during the perinatal period, although poor measures of support and confounding variables such as psychosocial
factors which contribute to a woman's ability to adapt to her pregnancy, make interpretation difficult.

Social support during the perinatal period has been shown to be beneficial. A systematic review of twelve randomized trials found that providing support to women during the intrapartum period resulted in reduced likelihood of Cesarean delivery, use of intrapartum analgesia/anaesthesia, operative vaginal delivery, and five minute Apgar scores of less than seven (Hodnett, 1998). Another systematic review revealed that during the postpartum period, home-based social support for socially disadvantaged mothers and babies was associated with reduced hospital admissions for childhood injury, lowered use of the hospital emergency department, and fewer children with incomplete well-child immunization at two years of age. Mothers' reports of unhappiness were also reduced (Hodnett & Roberts, 1997).

**Summary**

In summary, postpartum depression is a multidimensional condition with conflicting reports regarding its incidence, duration, and origin. Although much speculation exists about the cause of postpartum depression, marital conflict or lack of support in the spousal relationship has been strongly associated with the development of postpartum depression. Practitioners are seeking an appropriate method of treatment for postpartum depression, but research is limited and the quality of available studies is poor. Although medication may be useful, social support may provide an alternative, complementary approach to treatment. A systematic review and meta-analysis was undertaken to identify and assess the randomized, controlled trials done through October, 1996, to evaluate the effectiveness of
professional/social support in the treatment of postpartum depression. Although a broad definition of social support was used in the search strategy, treatment by a professional support provider was the only research available. The next chapter presents the published review (Ray & Hodnett, 1997).
CHAPTER II  THE COCHRANE REVIEW

The following outlines the process used in the development of this review:

1. The title was registered with the Pregnancy and Childbirth Group, Cochrane Library, January 1997.

2. A protocol was developed according to approved Cochrane Collaboration procedure. Appendix B describes the process of protocol development in Cochrane Reviews. The protocol was sent to the Pregnancy and Childbirth Editorial office in Liverpool, U.K., for appraisal by an editor and reviewers. Comments on the protocol were returned to Karen L. Ray. Following revision, the protocol was accepted for publication in the Cochrane Library, June, 1997.

3. Studies were identified using search strategies outlined in the Cochrane Handbook and were obtained through the Pregnancy and Childbirth Specialty Trials Register and Index Medicus. A review of the identified studies was done using methods outlined by the Cochrane Handbook (Appendix C). Data were extracted from the studies and analyzed using meta-analysis. Meta-analysis is used to obtain an objective estimate of the results of treatment. Meta-analysis goes beyond critique and integration of research findings to conduct statistical analysis on the outcomes of similar studies (Burnes & Groves, 1993). The protocol defined acceptable outcomes for this review as maternal, infant or family morbidity. Both Holden et al. (1989) and Appleby et al. (1997) used the study
outcome of continued depression at 25 weeks postpartum. The data consisted of combined scores obtained on several depression scales. The dichotomous scores were compared between the group that received the intervention and the control group.

Following entry of the data, the heterogeneity of the studies was investigated. The confidence intervals for both studies were observed to overlap. This indicated that the differences between the two studies was not significant and their results could be pooled using a fixed effect model.

A fixed effect model is an appropriate test of significance for dichotomous data. The random effects model was not appropriate for these data because it relies on the assumptions that the studies are a random sample from a hypothetical population of studies and that the heterogeneity between studies can be represented by a single variance. Moreover, a pragmatic concern is that the random effects method gives more weight to smaller studies than the fixed effect model, and smaller studies are often of poorer quality and may be more susceptible to publication bias (Cochrane Library).

As the samples compared in the final analysis were similar to those found in a case-control design, the odds ratio was used as a summary statistic. Case-control studies are used in situations in which the likelihood of developing a condition is low. In these circumstances researchers assemble a group of people with the condition and an appropriate set of people without the condition and compare them. The odds ratio is computed using a Chi-square and the formula Odds Ratio= 
AD/BC (Streiner & Norman, 1996).

Interpretation of the results was done following the statistical analysis of the data. Although the odds ratio indicates a possible benefit to the treatment, clinical and research recommendations were made with caution due to the small sample sizes and wide confidence intervals.

4. The review was completed and sent to the above editorial office for external review. Comments suggesting editorial changes were sent to the first author and the revised review was accepted October, 1997. The review first appeared in Issue 1 of the 1998 edition of the Cochrane Library.

The full reference for the Review is:

Caregiver support for postpartum depression

Caregiver support for postpartum depression
Ray KL, Hodnett ED

Date of most recent substantive amendment: 23 September 1997

Objectives: To obtain unbiased estimates of the effectiveness of professional and/or social support interventions in the treatment of postpartum depression.

Search strategy: The search employed strategies developed by the Cochrane Pregnancy and Childbirth Group. Relevant trials were identified in the Group's Specialized Register of Controlled Trials.

Selection criteria: Studies were included in the review if they met the following criteria:
The purpose of the study was clearly defined as treatment of postpartum depression. Additional support from caregivers was compared with usual forms of care in the postpartum period. Assignment to study groups was random or quasi-random, and the methodological quality was acceptable, with no evidence of systematic error in the randomization process or patient follow-up. Trials were not excluded if participants also received pharmacological treatment.

Data collection and analysis: Data were obtained from published articles and from unpublished information obtained from one of the trial authors. Both reviewers extracted the data. For one trial, the number of participants originally allocated to each treatment group was obtained, in order to perform an 'intention to treat' analysis. The odds ratio and 95% confidence intervals were calculated for comparable categorical data. Results were pooled using a fixed effects model.

Main results: Outcomes are given as dichotomous depression scores at twenty-five weeks postpartum. Treatment of postpartum depression with support may be beneficial, but the results must be viewed with caution due to the small numbers of women enrolled in the trials.

Conclusions: Depression among new mothers is a significant health problem. Postpartum depression may be alleviated by enhanced professional and/or social support, but further studies are needed.

Background
Postpartum depression has become a broad, illness-dominated definition for unhappiness following childbirth (Romito 1989). Complex factors have contributed to varied reported rates and a lack of consensus on definition, symptoms and causes of postpartum depression. Theoretical and measurement issues further complicate its investigation and may contribute to the over- or under-identification of this problem.

Estimates of its incidence range from seven to thirty percent, depending on the criteria and instruments used in its assessment. In addition, duration has been poorly documented due to a lack of longitudinal studies (Romito 1989). Women experiencing postpartum depression appear to suffer disabling symptoms including excessive fatigue, insomnia, inability to cope and suicidal ideation (Affonso 1984; Pitt 1968). This morbidity may have a profound effect on the mother, infant and family (Cogill et al 1986; Murray 1992; Oates 1994).

Controversy exists as to the causes of postpartum depression. The medical model proposes that depression results from changes in hormone levels during the postpartum period, while personality and psychoanalytic theories suggest that pre-existing factors in the women's personality and history make her vulnerable to postpartum depression. These theories remain
unproven and ignore social factors, despite little evidence that postpartum depression exists in non-industrialized societies (Romito 1989).

The concept of social support is rooted in the belief that individuals need a meaningful relationship with others. Crucial aspects of this relationship are the perception or the experience of feeling that one matters to others (Stemp et al 1986). Social support appears to be beneficial during the perinatal period. Supportive relationships during the perinatal period may enhance feelings of well-being, personal control and positive affect, and thereby help women to perceive pregnancy-related changes as less stressful (Norbeck 1989). Support to women during labour has been shown to reduce length of labour, operative delivery and other medical interventions (see Cochrane Review 'Support from caregivers during childbirth'). During the postpartum period, home-based social support for socially disadvantaged mothers and babies has been associated with reduced hospital admissions, less use of the Emergency department and increased childhood immunization. Mothers' reports of unhappiness were also reduced (see Cochrane Review 'Home-based social support for socially disadvantaged mothers').

Although medication may be useful for the treatment of postpartum depression, compliance is often poor (Holden 1994). Social support may provide an alternative or complementary approach. The aim of this review is to determine whether there is evidence that professional and/or social support has a beneficial effect on the duration and resolution of postpartum depression.

Objectives
To obtain unbiased estimates of the effectiveness of professional and/or social support interventions for the treatment of postpartum depression.

Criteria for considering studies for this review

Types of participants
All primiparous or multiparous women and their partners who had a live infant at the time they were admitted into a trial and were identified as depressed postnatally were included. Therefore, women experiencing second trimester terminations were excluded. A broad definition of postpartum depression was used that excluded 'postpartum blues' and ignored the cause or timing of onset of the depression, to include women who were determined to be clinically depressed during the first six months postpartum.

Types of intervention
All types of professional and/or social support including emotional support, counselling, tangible assistance and information delivered by telephone, home or clinic visits, or individual or group sessions, compared with any form of usual care for depressed mothers.

Types of outcome measures
All unbiased estimates of the effectiveness of professional and/or social support interventions for the treatment of postpartum depression on indicators of maternal and family morbidity, including the duration and resolution of the depression and other indicators of social functioning.

Types of studies
 Studies were included in the review if they met the following criteria:

The purpose was clearly defined as the treatment, as opposed to the prevention, of postpartum depression. Additional support from caregivers was compared with usual forms of care in the postpartum period. Assignment to study groups was random, and the methodological quality was acceptable, with no evidence of systematic error in the randomization process or patient follow-up. Trials were not excluded if participants also received pharmacological treatment.

Search strategy for identification of studies
See: Collaborative Review Group search strategy

This review has drawn on the search strategy developed for the Pregnancy and Childbirth Group as a whole.

Relevant trials were identified in the Group's Specialised Register of Controlled Trials. See
Review Group's details for more information.

**Methods of the review**

Trials under consideration were assessed for appropriateness of inclusion and methodological quality without prior consideration of their results. Included trial data were processed as described in: Mulrow CD, Oxman AD (eds), Cochrane Collaboration Handbook [updated 1 March 1997]. In: The Cochrane Library [database on disk and CDROM]. The Cochrane Collaboration. Oxford: Update Software; 1996-. Updated quarterly.

**Selection of the Studies**
The trial reports were examined by both reviewers and any disagreement about inclusion/exclusion were discussed until 100% agreement was obtained.

**Assessment of Methodological Quality**
Details of randomization, concealment allocation, blinding and exclusions from data analyses were recorded and evaluated. Unbiased methods of randomization that were considered acceptable include random numbers generated by computer or sequentially numbered opaque sealed envelopes containing computer generated random allocation. The first rater (Karen Ray) assigned a rating to each trial, based on the quality categories described in the Cochrane Collaboration Handbook. Only trials in categories A or B were included in this review. The secondary reviewer (Ellen Hodnett) reviewed the methodology used and results obtained for the quality rating. Results were compared and differences were discussed until 100% agreement was obtained.

**Extraction of Data**
Both reviewers extracted the data. Unpublished data were obtained from one of the trial authors (Appleby). Wherever possible the number of participants originally allocated to each treatment group was identified to allow for intention to treat analysis. The outcome measure was the number of women who were considered to be depressed/not depressed at 25 weeks postpartum.

**Data Synthesis**
The Appleby [Appleby 1997] trial involved four study groups: 1) fluoxetine plus a single counselling session, 2) placebo plus a single counselling session, 3) fluoxetine plus six counselling sessions, and 4) placebo plus six counselling sessions. For the purpose of this review, data from groups one and two were combined as the control group and data from groups three and four were combined as the treatment group.

The odds ratio and 95% confidence intervals were calculated for categorical results of comparable trials. Results were pooled using a fixed effect model.

**Description of studies**
See table of 'Characteristics of included studies'.

**Methodological qualities of included studies**
Two trials were reviewed. Holden et al [Holden 1989] used a method of random allocation that involved the use of random numbers, although it is unclear whether the process was centrally controlled. The researchers used the Edinburgh Postnatal Depression Scale (EPDS), in the identification of postpartum depression, an instrument with good evidence of reliability and validity. No validity and reliability information was provided for the Goldberg psychiatric interview and the criteria used to identify postpartum depression with this instrument were not clearly defined. The health visitors and outcome assessor were blinded to treatment allocation, although the participants were not. Co-interventions in addition to the support intervention were identified at follow-up and did not appear to have been controlled for a priori. Five subjects were excluded from the analysis following randomization.

In the trial by Appleby et al [Appleby 1997], allocation was performed using computer generated random numbers, although as in the Holden et al [Holden 1989] study, it is unclear whether the process was centrally controlled. No information was provided about the validity and reliability of the clinical interview used to identify those with postpartum depression. The outcome assessor was blinded to treatment allocation, although the participants and psychologist providing the intervention were not. A single counsellor provided varying levels of support to all of the study groups.
Results
The common outcome for both trials was depression at 25 weeks postpartum, which was significantly reduced in the groups receiving additional support (typical OR = 0.34; 95% CI 0.17, 0.69).

Summary of analyses
MetaView: Tables and Figures

Discussion
Depression among new mothers is a significant health problem. The social isolation experienced by many mothers, particularly in Western cultures may be a contributing factor to postpartum depression (Romito 1989). Regardless of its etiology, postpartum depression may be alleviated by enhanced professional and/or social support. To date only 117 women with postpartum depression have been studied. In addition, the Appleby trial had a large drop-out rate (30%). The subjects that dropped out were similar to those that completed the trial in some respects, but not in others. This raises the question of selective loss to follow-up and may have introduced bias.

Conclusions
Implications for practice
While there is good theoretical justification for the potential benefits of social support in the resolution of postpartum depression, it would be premature to make practice recommendations based upon two small trials. Only 137 women have been studied to date.

Implications for research
As with other studies of support for childbearing women (see Cochrane Reviews 'Support from caregivers during childbirth' and 'Home-based social support for socially disadvantaged mothers'), questions remain about the relative benefits of social (for example, lay person) versus professional (health visitor, nurse, midwife) support. Questions also remain regarding whether support is better provided through individual home visits, telephone calls or group sessions. The appropriate timing and duration of these interventions remains undetermined.

Future research should examine the usefulness of social support in the prevention as well as treatment of postpartum depression. Trials should also evaluate the effects of support on outcomes such as reduction of symptoms, hospital admission rates and long-term maternal, infant and family well-being.

Potential conflict of interest
None known.

Acknowledgements
We are grateful to Professor Appleby for providing unpublished data.

Characteristics of included studies
Table: Characteristics of included studies

Characteristics of excluded studies
Study: Seeley 1995
Not a randomised controlled trial. Women who received a program of support for postnatal depression were compared to historical controls.

Study: Wickberg 1996
Alternate allocation was used to assign participants to study groups.

References
References to studies included in this review
Appleby 1997 (unpublished data only)
depression. BMJ 1997;314:932-936. [9465]

Holden 1989 *(published data only)*


*indicates the major publication for the study*

References to studies excluded from this review

Seeley 1995

Seeley S, Murray L, Cooper PJ. Health visitor intervention in postnatal depression. An evaluation of the outcome for mothers and babies. Unpublished manuscript.

Wickberg 1996


Additional references

Affonso 1984


Cogill et al 1986


Holden 1994


Murray 1992


Norbeck 1989


Oates 1994


Pitt 1968


Romito 1989


Stemp et al 1986

Stemp P, Turner RJ, Noh S. Psychological distress in the postpartum period:

Coversheet

Title
Caregiver support for postpartum depression

Short Title
Caregiver support for postpartum depression

Reviewer(s)
Ray KL, Hodnett ED

Date of most recent amendment: 25 February 1998

Date of most recent substantive amendment: 23 September 1997

This review should be cited as:

Contact address:
Ms Karen L Ray
19 Richdale Court
Etobicoke
Ontario
CANADA
M9B 5Y8
Telephone: +416 234 2007
Facsimile: +416 234 5391
E-mail:

For information on the editorial group see: Cochrane Pregnancy and Childbirth Group

Extramural sources of support to the review
- None on file

Intramural sources of support to the review
- University of Toronto CANADA

Comment, Reply and Editorial notes
Two new trial reports were added on 16 February 1998 to ‘References to excluded studies from this review’.

Keywords
HUMAN; FEMALE; DEPRESSION-POSTPARTUM / therapy; DEPRESSION-POSTPARTUM / etiology; CHILDBIRTH; INCIDENCE; HORMONES / blood; POSTPARTUM-PERIOD; HOME-NURSING;
CRG Code: HM-PREG
<table>
<thead>
<tr>
<th>Study</th>
<th>Methods</th>
<th>Characteristics of included studies</th>
<th>Interventions</th>
<th>Outcomes</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appetby 1997</td>
<td>Randomized controlled trial. Allocation by computer generated random numbers, although it is unclear whether the process was carried out in a randomized fashion. 57 women were randomized to treatment groups: 51 completed the trial. Participants were allocated to 1 of 4 treatment groups: fluoxetine plus 1 counselling session, fluoxetine plus 6 counselling sessions, placebo plus 1 counselling session and placebo plus 6 counselling sessions. Group means obtained on 3 assessment instruments before and after treatment were compared for both samples of 61 and 67 women. Outcome assessor was blinded to treatment allocation although participants and psychologist providing intervention were not.</td>
<td>503 women from 2 urban hospitals and their districts were identified as depressed by screening with Edinburgh Postnatal Depression Scale (EPDS) at 6 weeks postpartum. 408 women agreed to go on to further testing. 218 women were found not to be depressed using the revised psychiatric interview and research diagnostic criteria. Of the 188 depressed women eligible for enrolment, 191 did not wish to participate. The most common reason for refusal was reluctance to take medication, most often because they expected to improve without treatment. Those entering the trial were significantly younger (p&lt;0.001) than those that refused, with the mean age for participants being 25.4 years. 28 women were prometopausal, 50 multiparous, 18 had complicated deliveries; 15 had Caesarean sections. 28 women were single. History of depression was very common among the participants. 51 and 39 women had previous major or minor depressions respectively. 28 had a family history of depression.</td>
<td>The intervention was provided by 1 psychologist who received brief training in cognitive-behavioural therapy. The psychologist provided either 1 or 6 counselling sessions. Each counselling session was to offer reassurance and practical advice to help correct feelings of not coping, lack of enjoyable activities, lack of practical support and caring for older children.</td>
<td>The outcome was the existence of depression, according to scores on 3 measures (EPDS, the Clinical interview and the Hamilton Depression Scale) at 25 weeks postpartum.</td>
<td>For the purpose of this review, the 2 groups who received 1 counselling session were combined as a single control group, and the 2 groups who received 6 counselling sessions were combined as a single experimental group. Drop out rates were similar in the 4 groups. Drop outs were younger than subjects who completed the study (23.7 (SD 6.2) years versus 28.3 (5.1) years; t=2.06, P=0.04) and more likely to have an unemployed partner (x²=3.8, df=1, P=0.05) and to have had a planned pregnancy (x²=4.6, df=1, P=0.03), but the groups did not differ on initial psychiatric morbidity scores, employment, obstetric complications, party, family history, or personal history of depression, including postnatal depression.</td>
</tr>
<tr>
<td>Hodten 1989</td>
<td>Randomized controlled trial. Allocation using unspecified method involving random numbers. 55 women were randomized to treatment groups; analyses were performed on outcome data for 50 women. Heath visitors and outcome assessor were blinded to treatment allocation, although participants were not.</td>
<td>734 women at five health centers in Scotland were identified as depressed by Edinburgh Postnatal Depression Scale (EPDS) screening at 6 weeks postpartum. Those that scored above 13 were assessed by psychiatric interview and EPDS at 12 weeks postpartum. 80 women were identified, 5 women did not wish to participate, and a further 5 did not complete the trial after random allocation. No information is provided on those subjects that withdrew, mean age was 26 years. 9 mothers were delivered by Caesarean section and 5 had forceps deliveries, 47 were married and 3 were single. Based on the partner's occupation, 8 were from social class I and 2, 20 from social class III, and 24 were from social class IV and V. Social variables and history of major or minor depressive disorder were similarly distributed between the two groups.</td>
<td>The intervention was provided by 17 health visitors after 12 weeks postpartum. Health visitors were a group of women with similar educational backgrounds (12 were midwives, 2 had teaching experience). Each health visitor received an instruction manual and weekly training in non-directive counselling. The health visitors made weekly half hour visits to the treatment group for 8 successive weeks.</td>
<td>The outcome was the existence of depression, assessed by 2 measures (the EPDS and the Goldberg psychiatric interview) at 25 weeks postpartum. It is unclear whether these subjects that were classified as depressed scored as depressed on one or both measures.</td>
<td></td>
</tr>
</tbody>
</table>
Comparison: support from caregivers
depression at 25 weeks postpartum

<table>
<thead>
<tr>
<th>Study</th>
<th>Expt n/N</th>
<th>Ctrl n/N</th>
<th>Peto OR (95% CI Fixed)</th>
<th>Weight %</th>
<th>Peto OR (95% CI Fixed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appleby 1997</td>
<td>9 / 28</td>
<td>19 / 33</td>
<td></td>
<td>54.7</td>
<td>0.37 [0.13, 0.99]</td>
</tr>
<tr>
<td>Holden 1989</td>
<td>8 / 26</td>
<td>15 / 24</td>
<td></td>
<td>45.3</td>
<td>0.29 [0.10, 0.96]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>17 / 54</td>
<td>34 / 57</td>
<td></td>
<td>100.0</td>
<td>0.33 [0.16, 0.59]</td>
</tr>
</tbody>
</table>

Chi-square 0.10 (df=1) Z=2.96
CHAPTER III  CONCLUSIONS

Clinicians require the best available evidence to make rational health care decisions. The use of a systematic review helps to accumulate the available evidence on a defined topic to assist in decision making. In the past, reviews were often prepared by authorities, who had an intimate knowledge of a field, the participants, the problems that arise, and other partly subjective but extremely relevant considerations (Eysenck, 1994). Today, the use of a systematic review allows for a broader, objective perspective on treatment recommendations and helps to encompass a substantial variety of specific treatment regimes, types of patients, and outcomes. The systematic review helps to gain greater generalizability and precision by including all of the available evidence from trials that pertain to an issue (Thompson, 1994). The increase in statistical power achieved in a systematic review helps to provide more precise estimates of the effects on health care than those delivered by individual studies (Thompson & Pocock, 1991).

Systematic reviews also have limitations. Treatment recommendations are dependent on the quality of the studies done in the area. Poor quality studies provide unreliable information. As indicated in the Cochrane review, the studies identified in this systematic review had several methodological problems. Systematic reviews are limited to published studies alone. This may distort the results of a systematic review because positive results are more likely to be published than negative ones (Dickersin, Chan, Chalmers, Sacks & Smith, 1987). In addition, systematic reviews often include trials using only one study design.

To address some of these issues, the Cochrane Collaboration was formed to provide
systematic reviews on clinical topics and reduce bias from the review process. To accomplish this, the Cochrane Collaboration maintains content validity by using authorities in a field in the editorial process. Clear methods have been developed to assist the reviewer to establish protocols, develop search strategies and define criteria for assessing the quality of studies (Appendices B and C). Only randomized controlled trials are included in the reviews to minimize bias.

In a randomized controlled trial, the intervention effects are examined under controlled conditions that permit the manipulation of the treatment and the control of extraneous factors, so that changes in outcome can be directly linked to the intervention. The randomized controlled trial is characterized by random assignment of participants to the experimental groups and control of contextual or situational factors that may potentially confound the participants’ responses to treatment. This experimental control reduces the potential effects of confounding factors and increases the confidence in attributing the observed changes to the intervention. As a result, the validity of the inference about the effectiveness of the intervention is increased (Cook & Campbell, 1979).

Unfortunately, the randomized controlled study design may fail to provide specific information to guide practice. Results of an effectiveness study usually indicate that the intervention is or is not effective in producing some outcome, while holding extraneous factors constant. But the causal inferences regarding the intervention’s effectiveness are not necessarily relevant to everyday practice. In real-life situations, multiple factors interact in a complex network of causes and effects. Therefore, the effect of an intervention tested under experimental circumstances lacks the understanding of the interrelations among these factors
and their effects on the interventions and their outcomes (Sidani & Braden, 1998).

An alternative study design might allow for knowledge regarding subject’s self-selection into the type and amount of treatment, examining the influence of extraneous factors on treatment delivery and outcomes, assessing the implementation of the intervention and the processes mediating the intervention effects (Sidani & Braden, 1998). For example, in the Appleby et al. (1997) study, a large number of women chose not to participate in the trial after being identified as depressed. The substantial loss of 101 potential participants raises questions regarding the generalizability of the results.

In addition, clinicians require detailed knowledge of how an intervention works, at what optimal dosage, with which client population, under what conditions, to produce what outcomes (Sidani & Braden, 1998). This systematic review demonstrated a reduced rate of depression in the intervention groups indicating that supportive interventions may be of benefit in the treatment of postpartum depression. Unfortunately, the information provided by the researchers on what constitutes the supportive interventions is very limited.

Based on a social model, both empirical studies in the review provided supportive interventions to women identified as depressed in the postpartum period. In the Holden et al. (1989) trial, health visitors were used to visit women in the treatment group for eight successive weeks. At least half an hour was to be spent counselling, infant care being discussed separately. Counselling was based on Rogerian or non-directive counselling methods. The major assumption of this approach is that talking about your feelings to an empathic and non-judgmental professional will help people to take a more positive view of themselves and their lives. The counselling emphasized the importance of listening to clients
and encouraged them to make decisions based on their own judgement rather than advice. No information is provided on the type of care the control group received.

In the Appleby et al. (1997) trial, the counselling was derived from cognitive-behavior therapy and was designed to be delivered by non-specialists in mental health. Each session was structured to offer reassurance and practical advice on four areas of concern to depressed mothers: feelings of not coping, lack of enjoyable activities, lack of practical support, and caring for older children. In addition, the first session, which lasted one hour, allowed women time to describe their current circumstances and emotional state.

Both studies used half an hour of treatment time. Holden et al. (1989) used a duration of treatment of eight weeks, while Appleby et al. (1997) used six sessions. Although both studies used counselling as the supportive intervention, little information regarding what type of support women found helpful was included. As well, a greater understanding of how the supportive intervention was delivered and under what conditions it was used would have been useful.

Implicit in a systematic review is the belief that the outcome of the review is to improve clinical practice. Community care is currently in the process of making major changes in practice. With alterations in the reimbursement system, payers are demanding more of health care providers than ever before. Policy makers of health care expect increased acceptance of responsibilities, including high-quality research and the utilization of those findings to develop a scientific base for practice (Burns & Groves, 1993).

Currently, dissemination of research findings has increased through research conferences and publications; but a limited amount of this knowledge is being utilized (Burns
& Groves, 1993). The Cochrane Library aims to correct this problem by using an accessible, regularly updated, CD-ROM format, to provide recommendations for evidence-based practice. In the future, this may assist community health providers to plan care to individuals and develop programs to meet the needs of larger groups, such as women with postpartum depression.

In addition to clinical practice recommendations, information derived from research should also be used for theory development (Burns & Groves, 1993). A theory consists of a set of concepts that are defined and interrelated to present a view of a phenomenon. A theory is developed from a combination of personal experiences, research findings, and abstract thought processes. Findings from research may be used as a starting point, with the theory emerging as the theorist organizes the findings to best explain the empirical world. To further the theoretical knowledge regarding social support, determining how the supportive intervention made the mother feel less depressed should be examined in future studies.

Future research must also look at other maternal, infant and family outcomes. The outcome of continued depression at 25 weeks postpartum gives a very narrow view of the experience of postpartum depression for women and their families. Questions also remain regarding the benefits of social versus professional support. Whether support should be provided through home visits, telephone calls, or group sessions and the timing and duration of these interventions should also be investigated.

In conclusion, depression among mothers is a significant health problem. Social support has been shown to be associated with improved mortality and mental health. The Cochrane Review provided an acceptable method of obtaining a systematic review on support
for the treatment of postpartum depression. Postpartum depression may be alleviated by enhanced professional and/or social support, but the use of support for a multi-dimensional condition such as postpartum depression may need to be examined by additional systematic reviews of studies using other types of research designs. Further research should also include more information regarding methods to deliver supportive interventions, the mechanisms by which social support works, and additional relevant outcomes.
References


*Perspectives in Biological Medicine, 6*, 116-127.


Appendix A

OVERVIEW OF THE COCHRANE COLLABORATION

The following is an overview of the Cochrane Collaboration. More information can be obtained on the references used in the preparation of this summary, the Collaboration and their database by accessing the Cochrane Library [database on disk and CDROM]. The Cochrane Collaboration. Oxford: Update Software; 1996-. Updated quarterly.
Principles

In pursuing its aims, the Cochrane Collaboration is guided by six principles: collaboration, building on people's existing enthusiasm and interests, minimising duplication of effort, avoidance of bias, keeping up to date, and ensuring access.

Aims

The aims of the Cochrane Collaboration are to prepare and maintain systematic reviews of randomised controlled trials of the effects of health care, and of other evidence when appropriate, and to make this information readily available to decision-makers at all levels of health care systems. Although the primary focus of the Collaboration is on evidence relating to the provision of care provided within the health services, other interventions that may have an impact on health will also be considered within the Collaboration, provided there are interested individuals who are prepared to review the relevant evidence systematically.

Helping to meet the need: The Cochrane Collaboration

The Cochrane Collaboration has developed in response to Cochrane's call for systematic, up-to-date reviews of all relevant RCTs of health care. His hope that the methods used to prepare and maintain systematic reviews of RCTs in pregnancy and childbirth would be adopted and developed by other specialities was reflected in the decision to found a Centre named after him, as part of the Research and Development Programme established to support the National Health Service in England. In the months after the opening of the Centre at the end of 1992, people all over the world expressed strong support for its aims. The concept and the reality of the Cochrane Collaboration emerged naturally as a result of this global support for Cochrane's ideas. The organisation and development of the Collaboration is now guided by discussions at annual colloquia, involving participants from all over the world.

The task of the Cochrane Collaboration is to prepare, maintain and disseminate systematic, up-to-date reviews of RCTs of health care, and, when RCTs are not available, reviews of the most reliable evidence from other sources. Although a massive effort is required to build, maintain and disseminate the database of systematic reviews of health care which Cochrane envisaged, it has become clear that the collaborative spirit required to make efficient progress already exists. A willingness to collaborate with others is a fundamental prerequisite for getting to grips with Cochrane's agenda. Thus, although those contributing to the Collaboration are named in its electronically published output, the Cochrane Collaboration itself belongs to all of the contributors, collectively.

Experience of preparing and updating systematic reviews on the scale needed is still limited. Further, the time required to prepare valid reviews is often grossly underestimated. Lack of experience and time often forces good scientists to produce scientifically inadequate reviews. The key to the success of the Collaboration is thus to find means of harnessing the specific interests and enthusiasm of individuals who support the overall objectives of the Collaboration, and to find ways of providing the support of various kinds which they need to prepare and maintain systematic reviews.
Systematic, up-to-date reviews of RCTs of health care

During the 1980s, people began to respond to Cochrane's criticisms. Trialists themselves, for example, have collaborated in preparing systematic reviews of the results of RCTs in cancer, [R3876], [R3728], [R3877], RCTs of anti-platelet drugs for cardiovascular disease, [R3878] and RCTs of antibiotics for decontamination of the digestive tract of patients receiving intensive care. These global collaborative endeavours have yielded very important information for guiding treatment and future research, and they are yardsticks against which the quality of other systematic reviews of RCTs will continue to be judged. International collaboration has also led to the preparation of systematic reviews of all of the RCTs relevant to the care of women during pregnancy and childbirth [R3886], and of new-born infants [R3765].

Cochrane lived long enough to appreciate these developments. In 1987, the year before he died, he referred to the systematic review of RCTs of care during pregnancy and childbirth as "a real milestone in the history of randomised trials and in the evaluation of care", and suggested that other specialities should copy the methods used [R3879].

As Cochrane emphasised, systematic reviews of RCTs must be kept up to date to take account of new evidence. If this is not done, important effects of health care (good and bad) will not be identified promptly, and people using the health services will be ill-served as a result. In addition, without systematic, up-to-date reviews of previous research, plans for new research will not be well informed. As a result, researchers and funding bodies will miss promising leads, and embark on studies asking questions that have already been answered.

Arrangements already exist for updating collaborative reviews of RCTs in several areas of cancer and cardiovascular disease. Similarly, reviews of RCTs in pregnancy and childbirth have also been maintained and extended as new evidence has become available. Updated analyses have usually been published using traditional media (journals, books, drug bulletins, etc.), and they will continue to be published in these ways in future. Electronic media, however, offer a particularly appropriate means of modifying and disseminating systematic reviews of RCTs as new evidence accrues, as has been demonstrated with electronically published reviews of RCTs of care in pregnancy and childbirth [R3880].
The need for more reliable reviews of research evidence

Although Cochrane's ideas have received increasingly explicit support from a wide range of commentators, progress in applying his principles in practice has been very slow. In part, this reflects the fact that strong evidence about the effects of health care often threatens a variety of vested interests. A more basic problem, however, is that valid evidence about the effects of health care, even though it may have been published, is not readily accessible to those who want it for making decisions [R3873].

It is unreasonable to expect people such as clinicians, policy makers or patients who want reliable information about the effects of health care to unearth all the relevant evidence from reports of original research. These are far too numerous and too dispersed to be of practical use. Most people must rely on reviews of the primary research as a way of coping with the information overload confronting them. [R1426] Reviews thus occupy a key position in the chain which should link the results of research at one end, to improved outcomes of health care at the other.

Unfortunately, the quality of reviews leaves much to be desired. This is because most reviewers do not approach their task systematically, with a respect for scientific principles, [R1174], [R650], [R3697], in particular, the control of biases [R3700], [R3874], and random errors. [R3872], [R3708], [R3716]. For example, most reviews of evidence about the care of patients with myocardial infarction have not reflected the strong evidence that has emerged in systematic reviews of the relevant RCTs [R3672]. The poor quality of most reviews has meant that advice on some highly effective forms of health care has been delayed for many years, and that other forms of care have been recommended long after controlled research has shown them to be either ineffective or actually harmful.

Cochrane recognised that people who wanted to take more informed decisions in health care did not have ready access to reliable evidence. In 1979, he wrote:

"It is surely a great criticism of our profession that we have not organised a critical summary, by speciality or subspeciality, adapted periodically, of all relevant randomised controlled trials." [R3875]
Good decisions about health care rely on more than good reviews of the results of research

As the Cochrane Collaboration develops, it will make the results of research assessing the effects of health care more easily available to those who want to take better decisions. However, as Cochrane made clear in *Effectiveness and Efficiency*, reliable evidence about the effects of specific elements of health care, although essential for improving decisions about health care and research, is only part of what is needed for better decision-making.

If better decisions are to lead to improved health, then effective mechanisms are needed for implementing them efficiently. Forms of care that have been shown to do more good than harm should be encouraged, while those that do more harm than good need to be discarded. And the many forms of care which have unknown effects should, as far as possible, be provided only as part of research studies to find out whether they help or do harm.

In addition, if people are to receive care which is appropriate, then policy makers and decision takers - ranging from ministers of health to individual clinicians and patients - must consider people’s needs, the availability of resources, and priorities.

Careful assessment of the needs of people using the health services will remain of paramount importance. An episode during Cochrane’s years as a prisoner of war, when he was trying to care for a young Soviet prisoner who was dying, illustrates dramatically the challenges which will continue to face those trying to provide effective care:

"The ward was full, so I put him in my room as he was moribund and screaming and I did not want to wake the ward. I examined him. He had obvious gross bilateral cavitation and a severe pleural rub. I thought the latter was the cause of the pain and screaming. I had no morphia, just aspirin, which had no effect. I felt desperate. I knew very little Russian then and there was no-one in the ward who did. I finally instinctively sat down on the bed and took him in my arms, and the screaming stopped almost at once. He died peacefully in my arms a few hours later. It was not the pleurisy that caused the screaming, but loneliness. It was a wonderful education about the care of the dying. I was ashamed of my misdiagnosis and kept the story secret". [R3887]

With this anecdote Cochrane reminds us that decisions about how best to meet people’s needs are often complex, and that the non-specific effects of sensitively provided care are often very important in helping patients.

Undue reliance on evidence derived from formal investigations about the effects of specific elements of health care is inappropriate. Nevertheless, even if systematic reviews of well designed evaluations of the effects of care are not sufficient for improving policies and decisions in health care and research, they remain essential if these decisions are to become better informed.

During his last illness, Cochrane was pleased to discover a well designed randomised trial comparing the effects on quality of life of different ways of trying to provide supportive care for people who were dying and for their families. The Cochrane Collaboration is evolving to try to meet people’s needs for reliable, up-to-date information of this kind about the effects of all forms of health care.
3. ORGANISATION

This section of the Handbook introduces the current organisation we have created in order to accomplish our goals. It begins with a narrative summary that describes how a Cochrane Review is created, and how the various Cochrane entities help that happen, and then goes on to describe, each of these organisational bits in greater detail.

The Cochrane Collaboration is young and evolving rapidly. The challenge of achieving our goals is enormous, and requires hard work, enormous good will, and very great attention to communication among all those involved. Because it is unlikely that any single approach to the organisation, communication and co-ordination of the Collaboration's work will prove 'ideal', its structures and working arrangements are being kept under continued review. They are standing items on the agendas of successive annual Cochrane Colloquia so that, after discussions and suggestions made in the light of accumulating experience, we can agree on useful modifications and implement them.

As an introduction to how we are operating in February 1995, let's consider how Cochrane Reviews are created. It all starts with individuals (health professionals, methodologists, and consumers) who recognise that they share a common interest in an issue in the effectiveness of health care: What is the best treatment for stroke?; How can we minimise the harm from malaria?; What manoeuvres are most effective in helping people quit smoking? One or a few of these individuals take it on themselves to act as Facilitators in bringing the rest together, and begin their facilitation by contacting their nearest Cochrane Centre. The Centre helps them identify other potentially interested individuals from all over the world (because the Centres share a database of individuals who have expressed interest in the Collaboration) and helps the Facilitators plan and run an Exploratory Meeting involving everyone who is sufficiently committed to participate in it.

As an outgrowth of that or subsequent Exploratory Meetings, one or several teams take on, for the rest of their careers, the preparation and maintenance of a portfolio of systematic reviews of the effects of specific health care interventions for the prevention, treatment, and rehabilitation of specific conditions (e.g., myocardial infarction, premature labour, tobacco addiction). These teams (which, when formally registered by the Collaboration, are called Collaborative Review Groups) work out their own methods for organising and supporting their work, aided along the way with continuing advice and information provided by Cochrane Centres. This means identifying, through consensus, the members of the group who will be responsible for co-ordinating the group (called the Editorial Team), the operating methods they will employ in preparing Cochrane Reviews on specific topics, and the administrative arrangements (co-ordinated by an Administrator) that will help it all happen. It also involves licensing the Collaboration to disseminate their reviews by assigning (in exchange for a single, highly symbolic penny!) their reviews to the Collaboration as a whole (so that no journal or other publisher can claim exclusive copyright over it).

Each of the reviewers in the Collaborative Review Group is responsible (alone or with others) for developing protocols for their reviews, using this Handbook and advice from Cochrane Centres for guidance. When the protocol has been approved by the editorial team, they become officially registered as a Collaborative Review Group.

They then get down to the challenge of preparing their Review, and in doing so come in contact with, and benefit from, other Cochrane entities. In the tough task of tracking down all the randomised trials that are relevant, they once again work with Cochrane Centres to learn which pertinent sources already have been catalogued and which may require additional examination, perhaps even through the hand-searching of pertinent journals. Often they must carry out this hand-searching themselves, training
for which can be assisted by their affiliated Cochrane Centre. They also can benefit from the work of another Cochrane entity that is formed by individuals who share common interests in a grouping of health care consumers (such as the elderly) or a branch of health care (such as public health) or a broad category of health interventions (such as physiotherapy). These groups, called Fields, impinge on all the Collaborative Review Groups that deal with forms of health care in their fields of interest, suggesting perspectives and potential members that should be considered, and often carrying out extensive hand-searching of the pertinent journals, providing the results of their efforts to the registry of trials (the need for world-wide collaboration in this task is underscored by the recent documentation that a large proportion of the trials that were important to the Cochrane Stroke Group were published in Japanese).

As they carry out their review, the Editorial Team employs a series of rigorous methods to assemble, appraise, and often meta-analyse the trials that are relevant to their question. In doing so, they encounter the work of another Cochrane entity, called Methods Working Groups, created to organise and disseminate the work of a large number of methodologists who have come together to improve the validity and precision of systematic reviews. For example, Collaborative Review Groups benefit from Methods Working Groups who developed high-quality, uniform methods for hand-searching journals, and members from a number of Methods Working Groups have played major roles in the creation and maintenance of the Review Manager (or "RevMan") software that helps Editorial Teams organise, prepare, analyse, and present their systematic reviews.

When the individual Systematic Review has been completed (e.g., on the effects of heparin in ischaemic stroke), it will be added to other systematic reviews (e.g., of all interventions relevant to stroke) into a set of related Reviews called a Review Module. As part of a Module, the review is contributed together will all the other systematic reviews to form the Cochrane Database of Systematic Reviews, which is disseminated in several ways: on computer diskette, on Compact Disk, and through the Internet. Cochrane Reviewers may submit their reviews to any and as many journals or books as they desire, going through whatever additional peer review and modification (but no substantive change in data or conclusions) they negotiate with the journal's or book's editors (but pointing out that exclusive copyright cannot be assigned).

Important to the subsequent maintenance and updating of Cochrane Reviews will be the ability of those who read them to provide feedback to those who write them (e.g., Have any relevant trials been omitted? Should any other interpretations of the results be considered?). As this edition of the Handbook is coming out at the same time as the first disk issue of the Cochrane Database of Systematic Reviews (CDSR), we can report that we are seeking easy and efficient ways for readers of Cochrane Reviews to provide feedback about them to the Editorial Teams. Subsequent versions of the Handbook will describe the evolution and success of these efforts.

In this same vein, we are seeking ways for Cochrane Fields to offer implications on individual Cochrane Reviews from their perspectives (e.g., How ought primary care providers implement these results? How might consumers use this information to improve their care?). As these methods and the resulting commentaries evolve they will be evaluated and modified to make them as helpful as possible to the ultimate users of Cochrane Reviews.

We can now consider the Cochrane entities in greater detail:
Collaborative review groups

The front line contributors to the Cochrane Collaboration are the reviewers who have actually rolled up their sleeves to prepare and maintain systematic reviews. Each reviewer is a member of a collaborative review group, which consists of individuals sharing an interest in a particular topic (stroke, for example). Collaborative review groups grow out of exploratory meetings of people who have recognised that they share an interest in preparing and maintaining systematic reviews of RCTs relevant to a particular problem. Members of the review group seek funding and other support for their activities from whichever sources they consider appropriate. Each collaborative review group is co-ordinated by an editorial team. The editorial team is responsible for assembling an edited module containing the reviews prepared by members of the review group for incorporation in and then dissemination through the 'Cochrane Database of Systematic Reviews' as well as other information about them.

The characteristics and activities of a particular collaborative review group - the Cochrane Pregnancy and Childbirth Group - help to illustrate how these principles work in practice. The group comprises over 30 reviewers and an editorial team of six people. Collectively, the group is currently responsible for maintaining about 600 systematic reviews of RCTs, and for dealing with between 200 and 300 new reports of trials every year. The group includes reviewers in Australia, Canada, Ireland, the Netherlands, South Africa, the United Kingdom and Zimbabwe. Each of these reviewers is responsible for obtaining the resources (particularly their time) needed to prepare and maintain the reviews that fall within their respective areas of expertise. An editorial team (supported by a grant from the Department of Health for England) consisting of four editors, an administrator and an administrative secretary co-ordinates the work of the collaborative review group.

The Pregnancy and Childbirth Group is responsible for preparing an edited module - the Pregnancy and Childbirth Module - for incorporation in and then dissemination as part of the main 'Cochrane Database of Systematic Reviews'. Members of the Pregnancy and Childbirth Group also use their electronically published reviews as a basis for preparing printed articles and books [R3882].

The key principles guiding the formation of collaborative review groups must be people's existing interests and enthusiasm and their preparedness to collaborate with others sharing a common interest. This said, a framework is needed to guide and co-ordinate the development of collaborative review groups, and this must minimise unnecessary duplication of effort and help ensure the broadest possible coverage of all forms of health care.

Most collaborative review groups should focus on specific health problems, defined as conditions that impair people's health or increase their risk for impaired health. Examples include stroke, myocardial infarction, diabetes mellitus, schizophrenia, and pregnancy. These health care problems should be considered "comprehensively" by collaborative review groups, to include consideration of all relevant aspects of care (prevention, treatment and rehabilitation) for all types of patients or people (young and old, men and women, rich and poor, developed and developing countries, etc.) who may experience the health problem in question.

To achieve this comprehensiveness of approach, review groups will need to be multidisciplinary, and involve reviewers and editors with a range of relevant backgrounds and interests (for example, medical, surgical, physiotherapy, developing country, occupational therapy, nursing, complementary therapies, primary health care, care of the elderly, paediatrics, etc.).

Some collaborative review groups may be needed to deal with management problems that are common
to a range of health problems, for example, the organisation of health services. Those proposing collaborative review groups to deal with management problems must make a convincing case that their reviews will not duplicate unnecessarily reviews focusing on health problems. When there is necessary or desirable redundancy, the relevant review groups should be aware of each others' work, keep redundancy to a minimum, remain in contact, and attempt to reach agreement on evidence and interpretation.

People working in clinical fields that are defined by a focus on the care of only some types of people (for example, children, the elderly, or people receiving primary health care) should form collaborative review groups only for those health care problems that fall solely within the field in question (for example, child development). The perspectives and interests of these fields should be fostered through field co-ordination in the variety of ways described in section III.
Cochrane centres

Cochrane centres share a responsibility for helping to co-ordinate and support the Cochrane Collaboration. There are eight centres at the time of writing with more likely to become established during 1995, and initial enquiries about the Collaboration should be communicated through one of these (contact details are shown at the beginning of the Handbook). The shared responsibilities of the Cochrane Centres include:

- maintaining a directory of people who have expressed interest in contributing to the Cochrane Collaboration, with information about their specific interests
- helping to establish collaborative review groups, by fostering international collaboration among people with similar interests, participating in exploratory discussions and meetings, helping to organise workshops, and in other ways, facilitating collaboration
- co-ordinating the Collaboration’s contributions to the creation and maintenance of an international register of completed and ongoing RCTs, thus facilitating the first phase of data collection for reviewers
- preparing and developing protocols and software - compiled in successive editions of the Collaboration's 'Tool Kit' - to systematise and facilitate the preparation and updating of systematic reviews
- making arrangements for efficient electronic transfer of reviews between reviewers and editors; between editors and the parent database of Cochrane Reviews, and between this parent database and the 'Cochrane Database of Systematic Reviews'.
- developing policies and setting standards to maximise the reliability of information disseminated through the 'Cochrane Database of Systematic Reviews'
- promoting and undertaking research to improve the quality of systematic reviews
- exploring ways of helping the public, health service providers and purchasers, policy makers and the press to make full use of Cochrane Reviews
- organising workshops, seminars and Colloquia to support and guide the development of the Cochrane Collaboration

Proposals for the formation of new Cochrane Centres should be sent to the Steering Group, who will use the following criteria in assessing proposals and "recognising" new Centres:

- they should have demonstrated both an enthusiasm for, and a basic expertise in, performing systematic reviews.
- they should be able to provide local help and training in performing systematic reviews.
- there should be evidence of support from the other relevant individuals and groups in their region.
- they should have an advisory board, including consumer involvement.
- there should be good prospects for their becoming self-supporting.
- reflecting the Cochrane spirit of unselfish collaboration, they should be willing to take on some task of service to the entire Collaboration.
Appendix B

SEARCH STRATEGIES

Search strategy for specialised register
THE COCHRANE PREGNANCY AND CHILDBIRTH GROUP’S SPECIALISED REGISTER OF
CONTROLLED TRIALS

1. INTRODUCTION

The editorial team of the Cochrane Pregnancy and Childbirth Group assembles, maintains and
administers centrally a specialised register of controlled trials as a service to the 73 reviewers
who are members of the Group. The basis for the register was established between 1978 and
1985 with the creation of the Oxford Database of Perinatal Trials (ODPT). The development
and features of ODPT have been described in a series of reports published over the past 15
years (see references, below).

The specialised register created to assist members of the Cochrane Pregnancy and Childbirth
Group currently contains more than 6000 records referring to completed or ongoing trials, with
an annual accrual rate of about 300 new records. Photocopies are obtained of all reports
identified, whether by the formal search methods described below, or by informal discovery.

On the basis of the health topic(s) and/or form(s) of care covered, every record in the register
has been assigned by the editorial team to one or more reviewers in the Cochrane Group,
according to the agreed spheres of responsibility of each.

2. ELIGIBILITY CRITERIA

TOPIC SCOPE: Controlled trials comparing alternative forms of care used either during
pregnancy (but not to terminate early pregnancy), or within 28 days of delivery.

STUDY DESIGN: A controlled trial has been defined as a trial involving humans in which
allocation to the intervention has either been at random, or by some quasi-random method,
such as by alternation, or on the basis of the case record number or date of birth.

These criteria have been applied fairly liberally to avoid excluding potentially useful studies
involving concurrent comparisons of alternative policies. In other words, the register includes
reports which, if necessary, can subsequently be rejected as methodologically inadequate by a
member of the Group preparing a systematic review.

3. SEARCH FOR ELIGIBLE CONTROLLED TRIALS

(i) Electronic search of bibliographic databases

The National Library of Medicine MEDLINE database has been searched back to 1966, and is
updated quarterly. The method of access and search strategy have been adjusted from time to
time. The current search strategy, using SilverPlatter MEDLINE, is as follows:

1 RANDOMIZED-CONTROLLED TRIAL in PT
2 RANDOMIZED-CONTROLLED-TRIALS
3 RANDOM-ALLOCATION
4 DOUBLE-BLIND-METHOD
5 SINGLE-BLIND-METHOD
6 CLINICAL-TRIAL in PT
7 explode CLINICAL-TRIALS
8 (clin* near trial*) in TI
9 (clin* near trial*) in AB
10 (singl* or doubl* or trebl* or tripl*) near (blinde or mask*)
11 (#10 in TI) or (#10 in AB)
12 PLACEBOS
13 placebo* in TI
14 placebo* in AB
15 random* in TI
16 random* in AB
17 RESEARCH-DESIGN
18 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #11 or #12 or #13 or #14 or #15 or #16
or #17
19 explode FETUS
20 explode INFANT-NEWBORN
21 explode PREGNANCY
22 #19 or #20 or #21
23 #18 and #22
24 TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
25 #23 not #24
26 PT=CONTROLLED-CLINICAL-TRIAL
27 #18 or #26
28 #27 and #22
29 #28 not #24

(ii) Handsearch of journals

A systematic journal handsearch (for perinatal trials only) of the journals listed below has been carried out and maintained from the initial date of publication of each journal, or from 1950, whichever was the earlier. Maintenance of the search of some journals * (mainly paediatric and anaesthetic) was discontinued with the registration of other Cochrane entities during 1993:

Acta Anaesthesiol Scand (and supplements): From 1st issue and continuing
Acta Obstet Gynecol Scand (and supplements): From 1950 and continuing
Acta Paediatr Scand*: From 1st issue through 1993
Am J Clin Nutr: From 1st issue and continuing
Am J Dis Child*: From 1950 through 1993
Am J Obstet Gynecol: From 1950 and continuing
Anaesth Intens Care: From 1st issue and continuing
Anaesthesia: From 1950 and continuing
Anesth Analg: From 1st issue and continuing
Anesthesiology: From 1950 and continuing
Arch Dis Child*: From 1950 through 1993
Aust NZ J Obstet Gynaecol: From 1st issue and continuing
Birth: From 1st issue and continuing
BMJ: From 1950 and continuing
Br J Anaesth: From 1950 and continuing
Br J Obstet Gynaecol: From 1st issue and continuing
Can J Anaesth: From 1st issue and continuing
Can Med Assoc J*: From 1950 and continuing
Clin Pharmacol Ther: From 1st issue and continuing
Curr Med Res Opin*: From 1st issue through 1993
Dev Med Child Neurol*: From 1st issue through 1993
Early Hum Dev*: From 1st issue through 1993
Eur J Obstet Gynecol Reprod Biol: From 1st issue and continuing
Geburtshilfe Frauenheilkd: From 1950 and continuing
Gynecol Obstet Invest: From 1st issue and continuing
Int J Gynaecol Obstet (and supplements): From 1st issue and continuing
Int J Obstet Anesthesia: From October 1994 to October 1995
JAMA: From 1st issue and continuing
J Am Coll Surg: From 1950 and continuing
J Gynecol Obstet Biol Reprod (Paris): From 1st issue and continuing
J Int Med Res*: From 1st issue through 1993
J Nurs Midw: From 1st issue and continuing
J Obstet Gynaecol: From 1st issue and continuing
JOGN Nurs*: From 1st issue through 1993
J Pediatr*: From 1950 through 1993
J Pediatr Gastroenterol Nutr*: From 1st issue through 1993
J Perinat Med: From 1st issue and continuing
J Reprod Med: From 1st issue and continuing
Lancet: From 1950 and continuing
Med J Aust: From 1950 and continuing
Midwifery: From 1st issue and continuing
N Engl J Med: From 1950 and continuing
Nurse Res*: From 1st issue through 1993
NZ Med J: From 1950 and continuing
Obstet Gynecol: From 1st issue and continuing
Pediatr Res*: From 1st issue through 1993
Pediatrics*: From 1950 through 1993
Practitioner: From 1950 and continuing
Prostaglandins*: From 1st issue through 1993
Reg Anesth: From 1st issue and continuing
S Afr J Obstet Gynaecol*: From 1st issue through 1993
Surg Gynecol Obstet*: From 1950 through 1993
Ugeskr Laeger*: From 1950 through 1993
Z Geburtshilfe Perinatol: From 1st issue and continuing
Zentralbl Gynakol: From 1950 and continuing

(iii) Handsearch of conference proceedings

American College of Obstetricians and Gynecologists' Annual Meeting: 36th, 37th, 39th, 40th, 41st
Argentinian Congress of Perinatology: 3rd
Australian Perinatal Society: 14th
Birth Conference: 1st, 2nd, 3rd, 4th, 5th, 6th, 7th, 8th, 9th
British Congress of Obstetrics and Gynaecology: 23rd, 25th, 26th, 27th
British Paediatric Association Annual Meeting: 14th, 15th, 16th, 27th, 60th, 61st, 62nd, 63rd, 65th
European Congress of Allied Specialists in Maternal and Neonatal Care: 4th
European Congress of Obstetrical Anaesthesia and Analgesia: 1st
European Congress of Perinatal Medicine: 6th, 6th, 8th, 10th, 11th, 12th, 14th, 15th
European Congress on Prostaglandins in Reproduction: 1st, 2nd
European Congress on Ultrasound in Medicine and Biology: 6th
Federation of the Asia-Oceania Perinatal Societies' Congress: 6th
International Confederation of Midwives Triennial Congress: 24th
International Congress on Psychosomatic Medicine in Obstetrics and Gynaecology: 3rd, 5th
International Society for the Study of Hypertension in Pregnancy (IISHP) European Branch: 1st
Priorities in Perinatal Care in South Africa: 4th, 7th, 10th, 11th, 14th, 15th
Society of Obstetricians and Gynaecologists of Canada Annual Meeting: 49th
Society of Perinatal Obstetricians' (USA) Annual Meeting: 3rd, 6th, 7th, 8th, 9th, 10th
Society for Gynecologic Investigation (USA) Annual Program: 31st, 34th, 37th, 39th, 40th
World Congress of Perinatal Medicine: 1st, 2nd
World Congress of Gynecology and Obstetrics: 11th, 12th, 13th, 14th
World Congress of Hypertension in Pregnancy: 7th, 8th, 9th, 10th

(iv) Surveys to identify unpublished and ongoing trials

During the second half of 1986 and early 1987, letters were sent to approximately 42,000 obstetricians and pediatricians in 18 countries in an attempt to identify unpublished controlled trials in perinatal medicine. The countries included in the survey were selected because they had generated more than 90% of the published reports of controlled trials in the Oxford Database of Perinatal Trials. This resulted in the notification of 395 unpublished randomized trials. Only 18 of the trials had been completed more than 2 years before the survey, a period during which at least 2300 reports of perinatal trials had been published. Of the 395 unpublished trials, 125 had ceased recruitment within the 2 years prior to the survey, 193 were actively recruiting at the time of the survey, and 59 were about to begin recruitment.

In 1991, prompted by the disappointing response to the earlier survey of individuals in an attempt to obtain information about unpublished and ongoing trials, a further, more focussed survey was conducted of clinical and academic institutions and funding agencies in the United Kingdom and North America to assess the feasibility of voluntary registration of trials. The experience gained in this and the earlier survey suggested that publication bias could not be addressed successfully by attempts to obtain information about unpublished trials retrospectively. This has led members of the Cochrane Pregnancy and Childbirth Group to support calls for prospective registration of trials, at inception.

4. REFERENCES

General references describing the creation and operation of the Cochrane Pregnancy and Childbirth Group's specialised register.

Chalmers I, Hetherington J, Elbourne D, Keirse MJNC, Enkin M. Materials and methods used


Publications describing identification and classification of published reports of trials:


References describing identification of unpublished and ongoing trials:


Appendix C

CRITICAL APPRAISAL OF STUDIES

The methods used in the review process are outlined in:

Mulrow CD, Oxman AD (eds), Cochrane Collaboration Handbook [updated 1 March 1997].
In: The Cochrane Library [database on disk and CDROM]. The Cochrane Collaboration.

6. CRITICAL APPRAISAL OF STUDIES

Critical appraisal of individual studies that are summarised in systematic reviews is necessary to limit bias in conducting the systematic review, gain insight into potential comparisons, and guide interpretation of findings. Parameters that warrant appraisal are those related to applicability of findings, validity of individual studies, and certain design characteristics that affect interpretation of results. Applicability, also called external validity or generalisability by some, is related to the definition of the key components of well-formulated questions outlined in section 4. Specifically, whether a review's findings are applicable to a particular population, intervention strategy or outcome is dependent upon the studies selected for review, and on how the studies as well as the reviewers define the people, interventions and outcomes of interest.

Interpretation of results is dependent upon the validity of the included studies and other characteristics. For example, a review may summarise twenty valid trials that evaluate the effects of antiischemic agents on symptoms of chest pain in adults with prior myocardial infarction. However, the trials may examine different preparations and doses of antiischemic agents and may have varying durations. These latter issues would affect interpretation though they may not be directly relevant to the internal validity of the trials. Examples of what and how to abstract data related to applicability and design factors likely to affect the interpretation of findings will be given in the next section (section 7). The remainder of this section will focus on critically appraising the validity of individual studies included in a systematic review. As most Cochrane Reviews focus on randomised trial data, we will concentrate on how to appraise the validity of such data.

6.1 Validity
6.2 Sources of bias in trials of healthcare interventions
6.3 Selection bias
6.4 Performance bias
6.5 Attrition bias
6.6 Detection bias
6.7 Approaches to summarising the validity of studies
6.8 Bias in non-experimental studies
6.9 Application of critical appraisal criteria
6.10 Incorporating assessments of study validity in reviews
6.11 Limitations of critical appraisal

References
6.1 Validity

In the context of a systematic review, the validity of a study is the extent to which its design and conduct are likely to prevent systematic errors, or bias \(240\). An important issue that should not be confused with validity is precision. Precision is a measure of the likelihood of random errors. It is reflected in the confidence interval around the estimate of effect from each study and the weight given to the results of each study when an overall estimate of effect or weighted average is derived using meta-analysis. More precise results are given more weight.

Variation in validity can explain variation in the results of the studies included in a systematic review. More rigorous studies may be more likely to yield results that are closer to the "truth". Quantitative analysis of results from trials of variable validity can result in "false positive" conclusions (erroneously concluding an intervention is effective) if the less rigorous studies are biased toward overestimating an intervention's effectiveness. They might also come to "false negative" conclusions (erroneously concluding no effect) if the less rigorous studies provide less precise or biased estimates of an intervention's effect \(135\).

It is important to systematically complete critical appraisal of all studies in a review even if there is no variability in either validity or results of the included studies. For instance, the results may be consistent among studies but all the studies may be flawed. In this case, the systematic review's conclusions would not be nearly as strong as if a series of rigorous studies yielded consistent results about an intervention's effect.
6.2 Sources of bias in trials of healthcare interventions
There are four sources of systematic errors in trials of the effects of healthcare: selection bias, performance bias, attrition bias and detection bias (see figure below). Unfortunately, we do not have strong empirical evidence of a relationship between trial outcomes and specific criteria or sets of criteria used to assess the risk of these biases (240, 491). There is, however, a logical basis for suspecting such relationships and good reason to assess these four potential biases (81).

**Sources of bias**
- **Selection bias** (systematic differences in comparison groups)
- **Performance bias** (systematic differences in care provided apart from the intervention being evaluated)
- **Attrition bias** (systematic differences in withdrawals from the trial)
- **Detection bias** (systematic differences in outcome assessment)

![Diagram of trial process]

**Target Population (baseline state)**

†

**Allocation**

‡ ‡

**Intervention Group**

†

**Control Group**

†

**Exposed to intervention**

†

**Not exposed to intervention**

†

**Follow-up**

†

**Follow-up**

†

**Outcomes**

†

**Outcomes**
6.3 Selection bias

One of the most important biases that may distort treatment comparisons is that which can result from the way that comparison groups are assembled (486). Using an appropriate method for preventing foreknowledge of treatment assignment is crucially important in trial design. When assessing a potential participant's eligibility for a trial, those who are recruiting participants and the participants themselves should remain unaware of the next assignment in the sequence until after the decision about eligibility has been made. Then, after assignment has been revealed, they should not be able to alter the assignment or the decision about eligibility. The ideal is for the process to be impervious to any influence by the individuals making the allocation. This will be most securely achieved if an assignment schedule generated using true randomisation is administered by someone who is not responsible for recruiting subjects, such as someone based in a central trial office or pharmacy. If such central randomisation cannot be organised, then other precautions are required to prevent manipulation of random assignment by those involved in recruitment.

The process of concealing assignment until treatment has been allocated has sometimes been referred to as "randomization blinding" (123). This term does not clearly distinguish concealed allocation from blinding of patients, providers, outcome evaluators and analysts and is unsatisfactory for three reasons. First, the reason for concealing the assignment schedule is to eliminate selection bias. In contrast, blinding (used after allocation of treatments) reduces performance and detection biases. Second, from a practical standpoint, concealing treatment assignment up to the point of assignment is always possible, regardless of the study question, but blinding after allocation may be impossible, as in trials comparing surgical with medical treatment. Third, control of selection bias is relevant to the trial as a whole, and thus to whatever outcomes are being compared. In contrast, control of detection bias is often outcome-specific and may be accomplished successfully for some outcomes in a trial but not others. Thus, blinding up to allocation and blinding after allocation of treatment are addressing different sources of bias, are inherently different in their practicability and may apply to different parts of a trial. To clearly distinguish these different forms and purposes of "blinding", we will refer to the process of concealing assignments as allocation concealment and reserve blinding for measures taken to reduce bias after treatment has been assigned.

Empirical research has shown that lack of adequate allocation concealment is associated with bias (123, 62). Indeed, concealment has been found to be more important in preventing bias than other components of allocation, such as the generation of the allocation sequence (e.g., computer, random number table, alternation). Thus, trials can be judged on the reported method of allocation concealment. Information should be presented that provides some assurance that allocations were not known until the point of allocation, at least. The method for assigning participants to treatments should be robust against patient and clinician bias and its description should be clear. The following are some approaches that can be used to assure adequate concealment schemes.

- centralised (e.g., group assignment by a central office unaware of subject characteristics) or pharmacy-controlled randomisation
- pre-numbered or coded identical containers which are administered serially to participants
- on-site computer system combined with group assignments in a locked unreadable computer file that can be accessed only after entering characteristics of an enrolled subject
- sequentially numbered, sealed, opaque envelopes
Other approaches may include statements that imply an approach similar to ones listed above, along with reassurance that the person who generated the allocation scheme did not administer it. Some schemes may be innovative and not fit any of the approaches above, but still seem to provide adequate concealment.

Approaches to allocation concealment that should be considered clearly inadequate include: alternation, case record numbers, dates of birth, day of the week, and any allocation procedure that is entirely transparent before assignment, such as an open list of random numbers of assignments. When trials do not report any concealment approach, adequacy should be considered unclear. Examples include merely stating that a list or table was used, only specifying that sealed envelopes were used and reporting an apparently adequate concealment scheme in combination with other information that leads the reviewer to be suspicious. When reviewers enter studies into Review Manager (RevMan) they are required to whether allocation concealment was adequate (A), unclear (B), inadequate (C), or that allocation concealment was not used (D) as a criterion to assess validity.
6.4 Performance bias

Performance bias refers to systematic differences in care provided to comparison groups other than the intervention of interest. To protect against unintended differences in care and placebo effects, those providing and receiving care can be "blinded" so that they do not know the group to which the recipients of care have been allocated. Some research suggests that such blinding is indeed important in protecting against bias (62, 487, 488). Studies have shown that contamination (provision of the intervention to the control group) and cointervention (provision of unintended additional care to either comparison group) can affect study results (489, 490). Furthermore, there is evidence that participants who are aware of their assignment status report more symptoms, leading to biased results (488). For these reasons, reviewers may want to consider the use of "blinding" as a criterion for validity. This can be done with the following questions: Were the recipients of care unaware of their assigned treatment? Were those providing care unaware of the assigned therapy?

A third question addressing blinding and detection bias is often added: Were persons responsible for outcome assessments unaware of the assigned therapy? This addresses detection bias, as noted below.

Blinding is likely to be particularly important in research with subjective outcome measures such as pain (62, 487, 488). Reviewers working on topics where blinding is likely to be important may want to develop specific criteria for judging the appropriateness of the method that was used for blinding. In some areas it may be desirable to use the same criterion across reviews, in which case a review group might want to agree to a standard approach for assessing blinding (69, 62, 480, 491).
Information from authors.

not apply to certain topic areas that have higher quality reporting or where it is possible to obtain missing
when it is implied rather than explicitly reported. This is a general recommendation, however, and may

Thus reviewers should be cautious about using reported follow-up as a validity criterion, particularly
implied in these reports on at-risk patients after ejaculation has not been found to be consistently related to bias [62].

biasing the results. It is also important to note that the mere reporting of inadequacies could itself be a problem. What is more likely is that reviewers should be cautious about implicit accounts of follow-up. The approach to handling losses has been described as handshaking protocols where losses of participants (e.g., withdrawal, death) are handled, reviewers should

pre-allocation exclusion and inclusion criteria for enrolment people. Because of inadequacies in reporting
bias has sometimes been referred to as exclusion bias, but we call it allocation bias to prevent confusion with

At this point, readers are systematic differences between groups in losses of participants from the study. It

6.5 Allocation bias
6.6 Detection bias

Detection bias refers to systematic differences in outcome assessment. Trials that blind outcome assessors regarding treatment allocation should logically be less likely to be biased than trials that do not. However, at least two studies have failed to demonstrate empirically a relationship between blinding of outcome assessment and study results, possibly due to inadequacies in trial reports (62, 245).

Somewhat different from bias in outcome assessment is bias due to selective reporting of results. This source of bias may be important in areas where multiple outcome measures are used, as in evaluations of treatments for rheumatoid arthritis (157). Therefore, reviewers may want to consider specification of predefined primary outcomes and analyses by the investigators indicators of validity. Alternatively, selective reporting of results could be taken to suggest the need for better reporting and efforts by reviewers to obtain missing data.
6.7 Approaches to summarising the validity of studies
6.7.1 Simple approaches
6.7.2 "Quality" scales and checklists
6.8 Bias in non-experimental studies

The logical reason for focusing on randomised controlled trials in Cochrane Reviews is that randomisation is the only means of allocation that controls for unknown and unmeasured confounders as well as those that are known and measured. Differences between comparison groups in prognosis, responsiveness to treatment or exposure to other factors that affect outcomes can distort the apparent magnitude of effects of the intervention of interest. It is possible to control or adjust for confounders that are known and measured in observational studies, such as case-control and cohort studies. However, it is not possible to adjust for those factors that are not known to be confounders or that were not measured. Unfortunately it can rarely, if ever, be assumed that all important factors relevant to prognosis and responsiveness to treatment are known, and for those that are known difficulties can arise in measuring and accounting for them in analyses. Empirical evidence supports these logical concerns (492). Selection bias can distort effects in either direction, causing them to appear either larger or smaller than they are. It is generally not possible to predict the magnitude, and often not even the direction of this bias in specific studies. However, on average, selection bias tends to make treatment effects appear larger than they are and the size of these distortions can be as large or larger than the size of the effects that are being measured (492).

Despite these concerns, there is sometimes good reason to rely on observational studies for information about the effects of healthcare interventions, and to include such studies in Cochrane Reviews. For example, well designed observational studies have provided useful data regarding the effects of interventions such as mandatory use of helmets by motorcyclists, screening for cervical cancer, dissemination of clinical practice guidelines to change professional practice and rare adverse effects of medication.

Various criteria have been suggested to critically appraise the validity of observational studies (493, 494, 495, 496). In general, the same four sources of bias noted above can be applied to other types of comparative studies, as illustrated below:

<table>
<thead>
<tr>
<th>Source of bias</th>
<th>Cohort studies</th>
<th>Case-control studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection bias</td>
<td>Control for confounders</td>
<td>Matching</td>
</tr>
<tr>
<td>Performance bias</td>
<td>Measurement of exposure</td>
<td>Measurement of exposure</td>
</tr>
<tr>
<td>Attrition bias</td>
<td>Completeness of follow-up</td>
<td>Completeness of follow-up</td>
</tr>
<tr>
<td>Detection bias</td>
<td>Blinding</td>
<td>Case definition</td>
</tr>
</tbody>
</table>

Concerns about attrition bias are similar in trials, cohort studies and case-control studies and relate to the extent that those entered into a study are appropriately accounted for in the results. Concerns about detection bias are also similar for cohort studies, and are related to the case definition that is used in case-control studies (since people are entered into such studies based on knowledge of the outcome of interest). The major difference between trials and observational studies has to do with selection bias and the need to identify and account for potential confounders in observational studies. To do this reviewers must make judgements about what confounders are important and the extent to which these were appropriately measured and controlled for. Assessing "performance bias" is also more difficult in
observational studies since it is necessary to measure exposure to the intervention of interest and ensure that there were not differences in exposure to other factors that could affect outcomes. In addition to considerations of blinding, which are similar to those in trials, it is important to consider whether exposure was measured in a similar and unbiased way in the groups being compared. So, for example, in addition to concerns about bias due to confounders in cohort and case control studies of the effects of post-menopausal hormone replacement therapy, investigators and reviewers must ensure that use of hormones was measured in an unbiased way.

In summary, a great deal of judgement is necessary in assessing the validity of observational studies. Judgement is also needed when the validity of trials is assessed, but the nature of observational studies makes them even more difficult to critically appraise. This requires a thorough understanding of both the problem that is the focus of the review and methodological considerations. Caution is advised.
6.9 Application of critical appraisal criteria

Several basic decisions must be made regarding the critical appraisal studies, similar to those made regarding the process of selecting studies (section 5.7). A prime consideration is the number of reviewers. Should there be one or more than one? How many are necessary and how many are too many? Will reviewers review the same articles to maximise reliability or mutually exclusive sets of reports to minimise workload? A concomitant consideration is reviewers' backgrounds and whether previous training and experience in study design or critical appraisal will be required.

Conducting systematic reviews with multiple reviewers is a two-sided coin. On the one hand it may limit bias and minimise errors and improve reliability of findings, but more than one creates the potential for disagreement among reviewers. When multiple reviewers are planned, there should be an explicit procedure or decision rule identified a priori for identifying and resolving disagreement. As a general rule, we recommend that at least two reviewers assess information that involves subjective interpretation and information that is critical to the interpretation of results (e.g., outcome data). The next section (section 7) describes methods for reaching and monitoring consensus when more than one reviewer is used.

Regardless of the number of reviewers, it is important to test any assessment criteria that are planned on a pilot sample of articles to ensure that the appraisal criteria can be applied consistently. A suggested sample would be three to six papers that span a range of low to high risk bias.

Should reviewers be especially trained in research methods, the content area of a review or both? Although experts in content areas may have pre-formed opinions that can bias their assessments, they may nonetheless give more consistent assessments of the validity of trials than persons without content expertise and they may have valuable insights that are different than those that someone with methodological expertise alone would have. It would seem intuitively desirable to use both content experts and non-experts and to ensure that both have an adequate understanding of the relevant methodological issues.

Reviewers must also decide whether those assessing study validity will be blinded to the names of the authors, institutions, journal and results of a trial when they apply critical appraisal criteria to the methods. Some empirical evidence suggests that blind assessment of reports might produce lower and more consistent scores than open assessments. However, such assessments are very time consuming. Reviewers must weigh the potential benefits of blind assessments against the costs involved in deciding whether or not to blind the reviewers. Further research is underway comparing blind and open assessments of trial validity and these results may help guide this decision.
6.10 Incorporating assessments of study validity in reviews

There are several ways in which validity assessments can be used in a review:

- as a threshold for inclusion of studies
- as a possible explanation for differences in results between trials
- in sensitivity analyses
- as weights in statistical analysis (meta-analysis) of the results

Failure to meet one or more validity criteria may indicate such a high risk of bias in some reviews that it constitutes grounds for exclusion of those studies. For example, for highly subjective outcomes such as pain, reviewers may decide to include only trials that prevent "performance bias" by blinding participants. The decision about where to set the cut point for inclusion can be conceptualised as existing on a continuum between "free from bias" and "undoubtedly biased" as illustrated below:

<table>
<thead>
<tr>
<th>&quot;free from bias&quot;</th>
<th>(variation in validity among included studies)</th>
</tr>
</thead>
<tbody>
<tr>
<td>included studies</td>
<td>cut-point</td>
</tr>
<tr>
<td>excluded studies</td>
<td>(e.g. all randomised controlled trials or all double blind randomised controlled trials)</td>
</tr>
</tbody>
</table>

"undoubtedly biased"

If reviewers raise the methodological cut-point for including studies, there will be less variation in validity among the included reports. Assessments of validity would then categorise studies by the risk of bias within the range above the cut-point for inclusion. With a sufficiently high cut-point, variation in validity among included reports may be moot.

There are several methods to examine whether validity may explain differences among study results. Visual plots of the results arranged in order of their validity can be used. A second approach is to analyse subgroups of studies above a methodological cut-point, preferably one specified \textit{a priori}. This approach can be used whether or not the results are heterogeneous, as a sensitivity analysis to determine if the overall results are the same when only studies with little risk of bias are included in the analysis. A third approach is to combine the results of each study sequentially in order of their assessed validity (a type of "cumulative meta-analysis"), examining the impact on the overall results as trials of decreasing validity are included (see section 8.6.2).
A fourth approach is to use statistical methods to weight studies according to their assessed validity or to use "meta-regression" to explore the relationship between validity and the magnitude of effect across studies (see section 8.6.1). Statistical methods for combining the results of studies generally weight the influence of each study by the inverse of the variance for the estimated measure of effect. In other words, studies with more precise results (narrower confidence intervals) are given more weight. It is also possible to weight studies according to validity so that more valid studies have more influence on the summary result. The main objection to this approach is that there is no empirical basis for determining how much weight to assign to different validity criteria or for quantitatively relating differences on any of the available "quality" scales to differences in the risk of bias.

It is possible using RevMan 3.0 to order studies according to either adequacy of concealment of allocation or "user defined" assessments of validity. Subgroup analyses based on assessments of validity can be done, although a test of statistical significance of differences between subgroups of studies has not been implemented. A function to facilitate sensitivity analyses also has not yet been implemented, but it is possible to do these by deleting or adding studies. RevMan does not include an option for weighting studies by methodological validity and neither cumulative meta-analysis or meta-regression is possible using RevMan 3.0.
6.11 Limitations of critical appraisal

There are two major difficulties with critically appraising the validity of studies. The first is inadequate reporting of trials (67, 213, 497). It is possible to assume if something was not reported it was not done. However, this is not necessarily correct. Reviewers should attempt to obtain additional clarifying data from investigators, but this may be difficult. The application of standards for reporting trials (67, 497) can facilitate critical appraisal.

The second limitation, which in part is a consequence of the first, is limited empirical evidence of a relationship between parameters thought to measure validity and actual trial outcomes. As noted above, there is empirical evidence suggesting that, on average, both inadequate concealment of allocation and lack of double blinding result in over-estimates of the effects of treatment. Clearly much more research needs to be done to establish which criteria for assessing validity are indeed important determinants of study results and when. Improved reporting of methods will also facilitate such research. Meanwhile, reviewers should avoid the use of "quality scores" and undue reliance on detailed quality assessments. It is not supported by empirical evidence, it can be time-consuming and it is potentially misleading.