UNIVERSITY OF STIRLING

Department of Psychology

The development and testing of an algorithm to support midwives' diagnosis of active labour in primiparous women

Helen Cheyne

A thesis submitted for the degree of Doctor of Philosophy

June 2008

ABSTRACT

The research in this thesis aimed to develop an algorithm to support midwives' diagnosis of active labour in primiparous women and to compare the effectiveness of the algorithm with standard care in terms of maternal and neonatal outcomes. Four linked studies are presented following the template suggested by the Medical Research Council (MRC 2000) Framework for development and evaluation of randomised controlled trials (RCT) for complex interventions to improve health.

Study one

Aim: To develop an algorithm for diagnosis of active labour in primiparous women.

Methods: An informal telephone survey was conducted with senior midwives to assess the need for a decision support tool for the diagnosis of active labour. A literature review identified the key cues for inclusion in the algorithm which was then drafted. Focus group interviews were conducted with midwives to ascertain the cues used by midwives in diagnosing active labour.

Findings: Thirteen midwives took part in focus groups. They described using informational cues which could be separated into two categories: those arising from the woman (*Physical signs*, *Distress and coping*, *Woman's expectations* and *Social factors*) and those from the institution (*Midwifery care*, *Organisational factors* and *Justifying actions*).

Study Two

Aim: Preliminary testing of the algorithm

Methods: Vignettes and questionnaires were used to test the consistency of midwives' judgements (inter-rater reliability), the content of the algorithm and its acceptability to midwives (face and content validity). The study was conducted in two stages: the first stage (23 midwives) involved vignettes and questionnaires and the second stage (20 midwives) involved vignettes only.

Findings: In the first stage a Kappa score of 0.45 indicated only moderate agreement between midwives using the algorithm. After modifying the algorithm, the Kappa score in stage two was 0.86, indicating a high level of agreement. While the majority of the midwives reported that the algorithm was easy to complete, most were able to identify snags or make suggestions for its improvement. Based on the findings of this study the algorithm was modified and the final version was developed.

Study three

Aim: To assess the feasibility of carrying out a cluster randomised trial (CRT) of the algorithm, in Scotland. Specifically, to identify maternity units potentially willing to participate in a CRT, to test the implementation strategy for the trial and to collect baseline data to inform the sample size calculation.

Methods: A questionnaire and interviews were used. The CRT methods were piloted in two maternity units and the algorithm was used for a three-month

period in order to test its acceptability and provide estimates of compliance and consent rates.

Results: All maternity units surveyed expressed an interest in the proposed study. Midwives' compliance with study protocol differed between units, although the consent rate of women was high (89% and 84%). Ultimately, one unit achieved 100% of the required sample and the other 60%. The midwives reported that the algorithm was acceptable and was a useful tool, particularly for teaching inexperienced midwives.

Study four

Aim: To compare the effectiveness of the algorithm for diagnosis of active labour in primiparous women with standard care in terms of maternal and neonatal outcomes.

Method: A cluster randomised trial

Participants: Fourteen maternity units in Scotland. Midwives in experimental sites used the algorithm to assist their diagnosis of active labour. Seven experimental units collected data from 1029 women at baseline and 896 post intervention. The seven control units had 1291 women at baseline and 1287 after study implementation.

Outcomes: The primary outcome was the percentage use of oxytocin for augmentation of labour. Secondary outcomes were medical interventions in

labour, labour admission management, unplanned out of hospital births and clinical outcomes for mothers and babies.

Results: There was no significant difference between groups in percentage use of oxytocin for augmentation of labour or for the use of medical interventions in labour. Women in the algorithm group were more likely to be discharged from the labour suite following their first labour assessment and subsequently have more pre-labour admissions.

Conclusion

The studies presented in this thesis represent the full process of developing and testing a complex healthcare intervention (the algorithm). The final study, a national cluster randomised trial, demonstrated that the use of the algorithm did not result in a reduction in the number of women who received oxytocin for augmentation or the use of medical interventions in labour. The results suggest that misdiagnosis of labour is not the main reason for higher rates of intervention experienced by women admitted to labour wards while not yet in active labour. These studies contribute significantly to the debate on care of women in early labour, the organisation of maternity care and to maternity care research.

ACKNOWLEDGEMENTS

It would not have been possible to successfully conduct this research without the contribution of a multidisciplinary research team, and because of this, my role in the research is specifically described.

The research presented in this thesis was developed, conducted and led by me, drawing as appropriate on the expertise and support of my research collaborators who formed a steering group for the CRT.

Research grants

The research was supported by two grants from the Scottish Executive Health Department Chief Scientist Office.

Grant 1. A feasibility study for a cluster randomised trial to investigate the use of a decision aid for the diagnosis of active labour in term pregnancy: The Early Labour Study in Scotland. (CZH/4/46). Grant holders: Helen Cheyne (Principal Investigator), Dawn Dowding, Vanora Hundley, Jill Mollison, Ian Greer.

Grant 2. A cluster randomised trial to investigate the use of a decision aid for the diagnosis of active labour in term pregnancy: The Early Labour Study in Scotland.

(CZH/4/245). Grant holders: Helen Cheyne (Principal Investigator), Vanora Hundley, Dawn Dowding, Paul McNamee, Catherine Niven, John Martin Bland, Lorna Aucott, Ian Greer.

All data for studies one and two were collected by me, data for studies three and four were collected by local study co-ordinators. Qualitative data analysis was conducted by me along with Dawn Dowding and Vanora Hundley. I developed the quantitative data analysis plan and conducted all the descriptive data analysis. The Cohen's Kappa statistics (study two) and regression analysis (study four) were conducted by the study statistician Martin Bland.

Publications arising from this research to date:

These publications are included at the end of the thesis (Appendix 6) with the permission of Wiley-Blackwell Publishing and Elsevier.

Cheyne, H., Dowding, D., Hundley, V., 2006. Making the diagnosis of labour: midwives' diagnostic judgement and management decisions. *Journal of Advanced Nursing*, 53 (6), 625-635. Copyright Wiley-Blackwell Publishing.

Cheyne, H., Dowding, D., Hundley, V., Aucott, L., Styles, M,. Mollison, J., Greer, I., Niven, C., 2008. The development and testing of an algorithm for diagnosis of active labour in primiparous women. *Midwifery*, 24 (2) 199-213. *(Epub* 2007). Copyright Elsevier.

Thanks

My heartfelt thanks go to Catherine Niven for her inspirational guidance and for providing me with every opportunity.

Thanks are due to Dawn Dowding who was my research programme leader and PhD supervisor during the early stages of the study and has continued to participate as a research collaborator.

I am deeply grateful to Vanora Hundley for her tireless work and guidance in supervising the writing of this thesis and to Roger Watt for his patient guidance and support.

I would like to express my thanks to research assistants Maggie Styles and Carol Barnett for their hard work during the feasibility study and the cluster randomised trial, to my research collaborators, in particular to study statistician, Martin Bland and to Karen Graham for her assistance in formatting this thesis.

This research would not have been possible without the support and enthusiasm of the many midwives who took part as local investigators, data collectors and participants. I am indebted to them and to the women who participated.

The study was based at the Nursing Midwifery and Allied Health Professions Research Unit at University of Stirling. My thanks go to my friends and colleagues for their support and encouragement.

My heartfelt appreciation goes to my sons Josh and Jonathan for their confidence in their mum and for maintaining my sense of humour, to my mother who is a wonderful role model and to Bill for his love and support.

LIST OF ABBREVIATIONS

AN	Antenatal
ARM	Artificial Rupture of Membranes
BBA	Born Before Arrival
CEMACH	Confidential Enquiry into Maternal and Child Health
CI	Confidence Interval
CONSORT	CONsolidated Standards of Reporting Trials
CRT	Cluster Randomised Trial
CS	Caesarean Section
CX	Cervix
DOH	Department of Health
DOMINO	Domiciliary In and Out
EFM	Electronic Fetal Monitoring
EGAMS	Expert Group on Acute Maternity Services
GP	General Practitioner
ICC	Intracluster Correlation Coefficient
ISD	Information and Statistics Division of the NHS Scotland
LDRP	Labour, Delivery, Recovery and Postnatal
MBU	Midwife Managed Birth Unit
MMPI	Minnesota Multiphasic Personality Inventory
MRC	Medical Research Council
MREC	Multi-centre Research Ethics Committee
NCT	National Childbirth Trust
NHS	National Health Service
NICE	National Institute for Clinical Excellence

NNU	Neonatal Unit
PI	Principal Investigator
RCM	Royal College of Midwives
RCOG	Royal College of Obstetricians and Gynaecologists
RCT	Randomised Controlled Trial
SD	Standard Deviation
SPCERH	Scottish Programme for Clinical Effectiveness in Reproductive Health
SRM	Spontaneous Rupture of Membranes
SVD	Spontaneous Vertex Delivery
UK	United Kingdom
US	United States
UTI	Urinary Tract Infection
VE	Vaginal Examination
WHO	World Health Organisation

CONTENTS PAGE

LIST OF TABLES AND FIGURES

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

Midwifery has a tradition that is woman centred and has as part of its discipline a foundation in science and also a foundation in art or intuition. As a midwife my research aims to contribute to the science of midwifery by providing evidence for midwifery practice while acknowledging the rich and multifaceted nature of midwifery practice.

Women are often uncertain about the onset of labour and in when to seek hospital admission; however this uncertainty may also extend to midwives and obstetricians. Although superficially straightforward, diagnosis of labour has been described as one of the most difficult and important elements in the care of a woman in labour (O'Driscoll et al. 1973; Lauzon and Hodnett 2000). This thesis suggests that introducing a decision support tool to assist clinician's diagnosis of labour has the potential to improve clinical outcomes for women.

1.1. Background

Throughout most of the last century government health policy advocated hospital birth (Tew 1990; Campbell and Macfarlane 1994), as a result the concept of home as the normal place of birth has become eroded. Planned home birth is now uncommon within the United Kingdom (UK) (between one and two percent of births), and across much of the developed world (US Department of Health and Human Services 1999; DOH 2005; ISD 2006), and

for a generation of women, hospital has become the traditional place to give birth.

This predominantly institutionalised model of care requires a clear cut, if somewhat artificial, distinction to be made between the latent phase of labour, a poorly defined period from onset of regular contractions, during which the woman might be expected to remain at home, and the active phase, the phase in which there is increasing cervical dilatation (Austin and Calderon 1999), when the majority of women would be admitted to hospital.

This distinction is important. Several studies have suggested that women who are admitted to labour wards early, that is, while not yet in labour, or while in the latent phase, are more likely to receive medical intervention during labour than women admitted in the active phase (Hemminki and Simukka 1986; Holmes et al. 2001; Jackson et al. 2003; Klein et al. 2003; Rahnama et al. 2006). The scale of the problem is illustrated by an audit of a workforce planning tool for midwifery services (Ball and Washbrook 1996) which reported that up to 30% of women admitted to labour wards in the UK were subsequently found not to be in labour.

It is not clear why women admitted to labour wards early receive more medical intervention. There may be factors intrinsic to the labours of some women which both lead them to seek early admission and to subsequently require

medical intervention. Alternatively, there may be factors involved in the admission itself, for example, it has been suggested that clinicians may misdiagnose active labour (Hemminki and Simukka 1986; Rahnama et al. 2006), or that their assessment of labour progress may be influenced by the length of time spent by a woman in labour ward (Hemminki and Simukka 1986) even where the woman is not yet in active labour.

Once a woman is admitted to labour ward the clock starts ticking (Simonda 2002; Kitzinger 2006) and her labour is generally expected to progress within strict time parameters; progress is usually monitored graphically by plotting cervical dilatation over time (Friedman 1989), with cervical dilatation of one centimetre per hour considered to be normal (WHO 1997). Where such progress does not occur a diagnosis of dystocia or 'slow progress of labour' may be made and the woman is likely to receive oxytocin to augment labour progress. A vignette-based study of doctors' decision-making demonstrated that simply varying the partogram information to make the labour appear longer, such as altering the ratio of time to cervical dilatation (flattening the curve) or including the latent phase of labour, encouraged doctors to intervene (Cartmill and Thornton 1992). Further, it has been suggested that there is a cascade effect of intervention in labour, where the use of one intervention triggers a series of further interventions ultimately contributing to increased use of operative or instrumental delivery (Inch 1985; Mold and Stein 1986; Tracy and Tracy 2003; Tracy et al. 2007).

The high and rising rate of seemingly routine intervention in labour has caused widespread concern (WHO 1997), as although appropriate medical intervention in labour has conferred health benefits on countless mother and babies, interventions have also been associated with increased morbidity and mortality. Reduction in the overall rate of intervention in labour is currently an international healthcare target (US Department of Health and Human Services 2000; CEMACH 2007). Good clinical judgement is essential in making the diagnosis of labour if unnecessary admissions and medical interventions in labour are to be reduced.

Diagnosis of labour would appear superficially to be a straightforward judgement and yet there is evidence that it is problematic in practice. Decision-making theory suggests why this may be the case through the concept of judgement under uncertainty (Tversky and Kahneman 1974; Hammond 1996) which suggests that many everyday judgements are made based on information that is unclear, 'noisy' and unpredictable. In addition, judgements may be made under time pressures and in an emotional atmosphere where there may be organisational and resource constraints and uncertainty of outcome. This has been described as judgement under conditions of 'irreducible uncertainty' (Dalgleish and de Michele 1995; Hammond 1996). Diagnosis of labour is one such judgement. Although the end point of labour is clearly defined with the delivery of the baby followed by the placenta and membranes (Crowther et al. 1991), there are many uncertainties surrounding the beginning of labour. Even fundamental aspects of labour onset, such as

when labour will start and what factors initiate labour, are not currently fully understood (McLean 2001; Smith 2001). In addition, the institutional birth setting which is currently predominant in many countries means that early labour assessment is frequently carried out in labour wards, or adjacent triage or assessment areas, which are characterised by time and workload pressures and an emotional atmosphere. In situations such as these, people are likely to make rapid intuitive judgements based on heuristics (mental short cuts) rather than analytical judgements. These types of judgements are prone to increased judgement error (Tversky and Kahneman 1974; Kahneman et al. 1982; Hammond 1996). The cognitive continuum theory suggests that it is the nature of the judgement task which determines the type of judgement style used and that altering aspects of the task, for example by applying a decision rule (such as an algorithm) which structures the judgement task, may induce a more rational judgement process thereby reducing error (Hamm 1988; Hammond 1996). Indeed, there is considerable evidence that use of decision support tools may improve clinical judgement (Grove et al. 2000; Dawes et al. 2002; Kawamoto et al. 2005). This thesis suggests that use of a decision support tool, in the form of an algorithm, to support midwives' diagnosis of active labour has the potential to reduce unnecessary labour ward admissions and improve clinical outcomes for women.

1.2. Research Aim

The aim of this research was to develop an algorithm to support midwives' diagnosis of active labour in primiparous women and to compare the

effectiveness of the algorithm, with standard care in terms of maternal and neonatal outcomes.

The research employed a mixed methods approach and was developed following the framework suggested by the Medical Research Council (MRC) (2000) for development and evaluation of randomised controlled trials (RCTs) for complex healthcare interventions. Diagnosis of labour may be considered superficially, not to be complex. However, the implementation of a decision support tool such as the algorithm is defined as a complex intervention for the reason that although it is targeted at the practice of the healthcare professional it is intended to have an impact on clinical outcomes. The MRC framework (2000) aims to improve the quality of trial design and implementation. It was chosen as a model for this research as it provides a template which is considered a gold standard for the development of trials of complex interventions in health care.

The framework suggests five phases in the development and implementation of a clinical trial, starting with consideration of the theoretical basis for the planned intervention, through paper based modelling and pilot phases, the exploratory trial, definitive RCT and finally consideration of possible long term implementation. The framework identifies important methodological issues to be considered at each phase for example, at the modelling and exploratory trial stages, the key importance of defining the intervention and of identifying the way in which the components of the trial will work together is highlighted (MRC

2000). This is essential both in developing and implementing the clinical trial and in the ultimate interpretation of the results.

This thesis comprises four linked studies, which broadly map onto the first four phases described in the MRC framework (2000) (Table 1), although not necessarily in sequential order. For example, the review of clinical and decision-making literature described in chapters two and three of the thesis maps onto the pre-clinical, theoretical development phase of the MRC framework, as does the development of trial methods described in chapter six.

Table 1 Study outline and mapping with MRC Framework (2000)

MRC Phase	Study	Aim	Thesis chapter
Pre-clinical Theory development		Clinical and decision-making literature reviews Strategic design development	2 & 3 6
Phase I Modelling	1 2	Development of the algorithm Preliminary testing of the algorithm	4 5
Phase II Exploratory trial	3	Feasibility study: to assess the feasibility of conducting a CRT of the use of the algorithm for the diagnosis of active labour in term pregnancy, in Scotland.	7
Phase III Definitive RCT	4	Cluster Randomised Trial: to compare the effectiveness of an algorithm for diagnosis of active labour, in healthy primiparous women, with standard care in terms of maternal and neonatal outcomes	6 & 8
Phase IV Long term implementation		Discussed	9

The final phase of the MRC framework (MRC 2000) considers long term implementation of the results of the study. This is beyond the scope of this thesis; however, the implications of longer term implementation of the results and further research required are discussed in chapter nine.

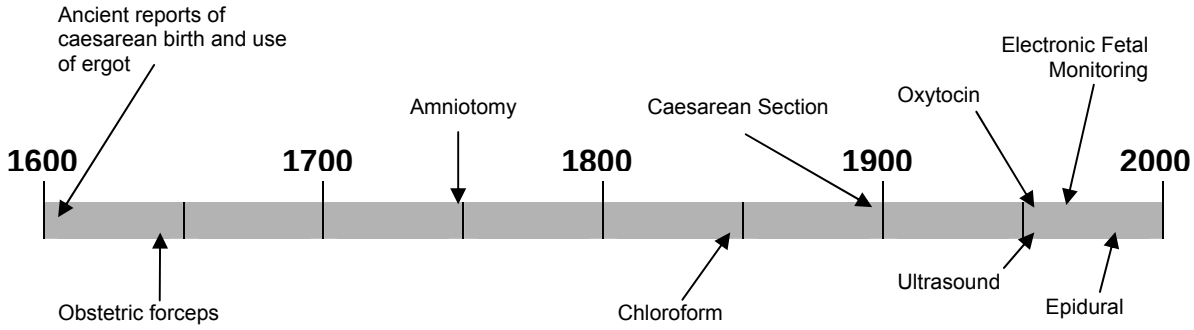
In the following two chapters the theoretical background of the research is presented. Chapter two reviews the literature with a clinical focus and considers the issues of increasing intervention in labour and aspects of labour diagnosis. Chapter three presents an overview of decision-making literature, in particular, rational and non-rational theories of judgement, and provides a theoretical framework for the choice of an algorithm as the study intervention.

CHAPTER 2: LITERATURE REVIEW

2.1. Intervention in Labour

Intervention in labour is not new and has been reported throughout history. For example, there are accounts of caesarean births in Roman times (although they were first successfully conducted in the UK at the end of the 19th century) (National Library of Medicine 1998) and ergot was used to control bleeding and stimulate labour for centuries (Mander 1998). Obstetric forceps were first used in the 17th century (Tew 1990), amniotomy in the 18th (Segal et al. 1999) and chloroform in 1847 (Mander 1998). However the practice of routine intervention in childbirth dates from around the middle of the 20th century when there was hot debate about the appropriateness of routine induction of post term labour (Kortenoever 1950; Wrigley 1958). The development of medical interventions in labour increased in pace from around that time (Figure 1), oxytocin for induction and augmentation of labour and obstetric ultrasound in the 1950s, electronic fetal monitoring in the 1960s and epidural analgesia in the 1970s (Tansey and Christie 2000; den Hertog et al. 2001; Martin 1998; Brill et al. 2003).

Figure 1 Development of medical intervention in labour



2.2. The increasing use of intervention in labour

It has been suggested that only a minority of women who give birth in the UK (or in many developed countries) do so without some form of intervention (Downe 2004). For example, a Department of Health report suggested that only 46% of women who gave birth in England (2003 to 2004) experienced normal birth, that is birth without surgical intervention, use of instruments, induction of labour, or augmentation of labour with oxytocin, epidural or general anaesthetic (DOH 2005). An Australian study of medical intervention in labour (Roberts et al. 2000) reported that only 18% of primiparous women receiving care in private hospitals between 1996 and 1997, achieved a vaginal birth without medical intervention, (39% of women attending public hospitals).

The key marker of the trend of increasing intervention is the rising rate of caesarean section, described by the World Health Organisation (WHO 2005) as an epidemic. In Scotland the rate increased from 9% in 1976 to 24% in 2005 (ISD 2006), similarly, in England the rate was 9% in 1980 rising to 23% by 2003 (DOH 2005). Both North and, in particular, South America have been reputed to lead this trend. In the United States the rate of caesarean birth increased from 23% in 1990 to 28% in 2003 (US Department of Health and Human Services 2005), while in Brazil the rate rose from 28% to 43% over a 20 year period (Costello and Osrin 2005) and the current rate is 80% amongst women who give birth in private sector hospitals (Potter et al. 2008). In Chile rates as high as 80% have also been reported (Murray 2000).

The use of different index interventions, denominators and the quality of data collection make comparison of other labour intervention rates difficult. However, it appears that these have also increased. A survey of electronic fetal monitoring in UK maternity units between 1985 and 1987 found that 63% of units monitored more than 60% of women (Wheble et al. 1989), while in a recent survey of women giving birth in England around 50% of women reported receiving continuous electronic fetal monitoring and 40% intermittent electronic fetal monitoring during labour (Redshaw et al. 2007), although there may have been overlap between these groups. A survey of maternity units in England conducted in 1984 reported that only 50% of maternity hospitals offered an epidural analgesia on request (Garcia and Garforth 1989), while data from 2003 to 2004 suggested that overall 21% of women giving birth in England had an epidural analgesia (DOH 2005). An exception to this trend was the UK rate of induction of labour which peaked at 48% in Scotland in 1976 before dropping to around 26% in 2005 (ISD 2006). There was a similar trend in England where the rate was 41% in 1974 and 20% in the period 2003 to 2004 (MacFarlane and Mugford 2000; DOH 2005).

Considering labour interventions worldwide, in 1989 more than two thirds of women in the US (68.4%) received electronic fetal monitoring; by 1997 this had risen to 83.3%. Over the same period induction of labour increased from 9.0% to 18.4% and augmentation of labour from 10.9% to 17.4% (US Department of Health and Human Services 1999). While in Brazil an increase from 3% to 45% was reported over 20 years (Costello and Osrin 2005). Intervention rates in

Australia are also high and the effect of private healthcare has been noted. Roberts et al. (2000) reported that 48.8% of primiparous women receiving private hospital care either had induction or augmentation of labour, 50.8% had an epidural and 16.4% had a caesarean birth.

2.3. Intervention in labour and maternal and fetal morbidity

There is ongoing debate about the appropriate use of intervention in labour (Wagner 1994; Johanson et al. 2002; WHO 1985; WHO 1997; Klein et al. 2006). Interventions now routinely used were developed in anticipation that they would confer benefit, and mothers and babies worldwide have benefited from their appropriate use, for example, oxytocin to prevent post partum haemorrhage or caesarean birth in cases of obstructed labour. The WHO estimates that without health care, including skilled professional care and use of appropriate medical intervention in birth, the maternal death rate worldwide would be four times the current level (WHO 2005). However, many interventions were widely adopted into clinical practice without adequate evaluation (WHO 1985; 1997; Chalmers 1992; Sandall 2004) and quickly became part of routine maternity care (for example, within 20 years of the development of oxytocin almost 50% of births in Scotland involved induction of labour (ISD 2006)). The use of some interventions has subsequently been associated with increased morbidity for mothers and babies. Electronic fetal monitoring is often used as an example of an intervention in childbirth which, having gained widespread acceptance, has had unintended consequences (Freeman 2007). It was introduced with the aim of identifying hypoxic babies so that treatment, in the form of expedited delivery could be performed; however

subsequent Cochrane Reviews (Thacker et al. 2006 (first published 1997); Alfirevic et al. 2006) have found that it did not reduce the perinatal death rate, and although associated with a 50% reduction in neonatal seizures, there was no evidence of long term benefit, specifically, a reduction in the incidence of cerebral palsy. Further, use of electronic fetal monitoring has been associated with a significant increase in caesarean birth and instrumental delivery (Alfirevic et al. 2006).

Instrumental and operative deliveries have been the focus of particular concern. MacArthur et al. (2001) in a study of primiparous women's health three months after delivery found that women who had a forceps delivery were nearly twice as likely to develop faecal incontinence. While this study (MacArthur et al. 2001) suggested that caesarean birth led to a slightly reduced likelihood of developing faecal incontinence, it may increase the risk of other types of morbidity for mothers and babies. Villar et al. (2007) reported that while caesarean birth had a protective effect for the baby in breech presentation it significantly increased the risk of severe maternal and neonatal morbidity and mortality. Hansen et al. (2008) found an increased risk of respiratory morbidity in babies born by elective caesarean section when compared to vaginal birth or emergency caesarean section, while Knight et al. (2008) reported that women having a caesarean section were at increased risk of peripartum hysterectomy, a risk which increased in women with previous caesarean births. Although much of the mortality and morbidity associated with caesarean birth may result from factors which lead up to the operative delivery, it appears that the

intervention itself confers increased risk. The confidential enquiry into maternal and child health (CEMACH 2007) highlighted the increased short and long-term maternal morbidity and mortality associated with caesarean section, in particular, from thromboembolism, haemorrhage, sepsis and anaesthesia and concluded (CEMACH 2007, p 84) that 'the operation is not as risk free as many have thought'.

Other interventions have also been linked with subsequent physical and psychological morbidity. For example, epidural analgesia has been associated with persistent backache and headaches and there is continuing debate about its contribution to increased rates of instrumental or operative delivery (Anim-Somuah et al. 2005; Klein 2006). Creedy et al. (2000) explored the incidence of acute trauma symptoms and posttraumatic stress disorder resulting from women's labour and birth experiences. This study found that women who had intervention in labour, in particular emergency caesarean section or forceps delivery, were at increased risk of suffering from acute trauma symptoms and that this risk increased if women were also dissatisfied with their care.

Finally, there are economic costs relating to use of intervention in labour. A study of the cost of different types of delivery (Petrou and Glazener 2002) reported significantly increased National Health Service costs associated with instrumental and operative delivery compared to spontaneous vaginal delivery, with operative delivery costing almost double that of spontaneous vaginal delivery. Tracy and Tracy (2003) conducted a study which aimed to estimate

the economic costs of a range of intrapartum interventions in addition to type of delivery. A cost model was developed in which four groups of labour interventions and possible birth outcomes were compared to spontaneous vaginal delivery with no intervention. The study found that for primiparous women costs increased with each additional intervention. Induction or augmentation of labour conferred an 11% cost increase, there was a further 20% increase associated with epidural analgesia and an additional 13% for women who received induction or augmentation as well as epidural.

The widespread routine use of medical intervention has become an issue of worldwide concern (Chalmers 1992; Wagner 1994; Johanson et al. 2002; Costello and Osrin 2005; Bick 2006; NCT/RCM/RCOG 2007). In a series of reports spanning twenty years the WHO has expressed concern about the inappropriate use of medical intervention in labour and identified interventions, commonly used which are of unproven benefit or harmful (WHO 1985; 1997; 2005), suggesting that '*the uncritical adoption of a range of unhelpful, untimely, inappropriate and/or unnecessary interventions, all too frequently poorly evaluated, is a risk run by many who try to improve the maternity services*' (WHO 1997, p1). In the US decreasing the rate of caesarean birth is a goal of the Healthy People Year 2000 and 2010 objectives (US Department of Health and Human Services 1990 and 2000), while within the UK maternity care guidelines have been produced with the aim of limiting unnecessary intervention in labour (NICE 2004 and 2007; RCM 2005).

2.4. Factors contributing to the routine use of labour intervention

A number of possible explanations have been suggested for increasing rates of intervention in labour, including protocols which make intervention based care the norm, fascination with gadgetry, commercial pressure; fear of litigation, women's choice, and even a failure by midwives to define normal labour (Wagner 1994; Gould 2000; Johanson et al. 2002; WHO 2005; Klein et al. 2006; Green and Baston 2007). All of these factors (as well as combinations) may have a contributing effect; however, two interesting trends have been noted. Women who have one intervention tend to receive a number of interventions (the cascade effect), and women who are admitted to labour wards early appear to receive more interventions.

2.4.1. The cascade effect

A number of authors have suggested that there is a cascade effect of intervention in labour (Inch 1985; Mold and Stein 1986; Hundley et al. 1994; Roberts et al. 2000; Tracy and Tracy 2003; Tracy et al. 2007). The notion is that when a woman receives a first intervention this triggers a series of subsequent interventions in a feed forward loop. Consequently, each intervention makes the next more likely, culminating in an increased incidence of instrumental or operative delivery. There is some evidence for such a cascade. Cochrane reviews have suggested that epidural analgesia (Anim-Somuah et al. 2005) and electronic monitoring (Alfirevic et al. 2006) are associated with increased rates of instrumental and operative delivery respectively, while the use of oxytocin for induction or augmentation necessitates the use of continuous electronic fetal monitoring (NICE 2007).

Roberts et al. (2000) studied the effect of increasing use of interventions in labour (induction or augmentation of labour and epidural analgesia) on subsequent birth outcome. In relation to primiparous women they reported significantly higher rates of assisted or operative delivery with increasing use of labour interventions. A population based study of low risk women giving birth in Australia during 2000 and 2001 (Tracy et al. 2007) aimed to determine the association between labour intervention and birth outcome. This study found that women who had induction or augmentation of labour were twice as likely to have a caesarean section and one and a half times more likely to have an instrumental delivery than women who had no labour intervention. Overall, 32.9% of primiparous women received induction or augmentation of labour combined with epidural analgesia. Of these women 70.3% subsequently had either assisted or operative birth, (36.7% and 33.5% respectively) compared to 13.5% among women who had no labour intervention (Tracy et al. 2007). The association of labour interventions and birth intervention is not in itself evidence of cause and effect, however, Roberts et al. (2000) also found that intervention rates were higher in women receiving care in private hospitals compared to those in public hospitals, suggesting that factors other than solely clinical need were involved.

2.4.2. Early admission and labour intervention

It is possible that hospital admission itself may be the trigger for a cascade of intervention. Several studies have suggested that women admitted to labour wards early receive more intervention in labour than those admitted in more

advanced labour (Hemminki and Simukka 1986; Holmes et al. 2001; Klein et al. 2003; Jackson et al. 2003; Rahnama et al. 2006).

In a retrospective study Hemminki and Simukka (1986) aimed to establish the relationship between the timing of hospital admission and the progress of labour and use of labour interventions in primiparous women admitted in spontaneous labour. Women were categorised as 'early comers' if they reported having regular contractions for less than four hours before admission and 'late comers' if they reported regular contractions for more than four hours. An intrinsic speed of labour was calculated for each woman based on cervical dilatation on admission in relation to the length of time they reported having had regular contractions. This meant that women could be classed as early comers who had either an intrinsically slow or fast labour. Likewise late comers could also either have an intrinsically slow or fast labour.

The study found that the although the mean length of labour (defined from the onset of regular contractions until delivery) was significantly longer for late comers than early comers, the mean time from admission to delivery was only slightly shorter (meaning that late comers laboured for longer at home). There was no significant difference in use of interventions in labour or type of delivery outcomes between women who were early or late comers, but when the intrinsic speed of labour was taken into account, women who were early comers with intrinsically slow labours spent a significantly longer period of time in labour ward between admission and delivery and had significantly higher

rates of intervention in labour (artificial rupture of membranes, electronic monitoring, pain relief, oxytocin and caesarean section) than any of the other groups. Thus, women who sought admission after a short period of contractions *and* had low cervical dilatation on admission spent the longest period of time in the labour ward and had the greatest number of medical interventions. The study could not determine if there were factors inherent in these women (fear or anxiety), which led them to seek early admission and were responsible for the increased rate of medical intervention or if the longer period of time spent in labour ward was responsible. The authors suggested that the clinician's assessment of labour progress may have focussed on the length of time which the women had spent in hospital rather than their overall duration of labour, thus spuriously diagnosing slow progress of labour.

Holmes et al. (2001) conducted a retrospective study which examined the relationship between the cervical dilatation at which women presented in labour and the likelihood of caesarean section, as well as the use of interventions in labour. The study included 3220 (both primiparous and multiparous) women who presented in spontaneous labour at term and who delivered within 36 hours of first presentation. Women were characterised as early comers if they presented with a cervical dilatation of less than four centimetres and late comers if they presented with cervical dilatation of four centimetres or more. Early comers (both primiparous and multiparous) had significantly longer labours (defined as time from onset of strong regular contractions until delivery), spent less time in labour at home, had a higher caesarean section

rate and a higher rate of oxytocin and epidural analgesia than women who were late comers. The study found no difference in clinical outcome for women who were discharged home following their initial presentation at the hospital.

Hemminki and Simukka (1986) and Holmes et al. (2001) used different definitions of early comers. Holmes et al. (2001) used cervical dilatation only; therefore all early comers were women who presented with a cervical dilatation of less than four centimetres, while Hemminki and Simukka (1986) used time from onset of contractions prior to admission to define early comers. The group of women they defined as slow labourers are more similar to the early comers of Holmes et al. (2001) as both these groups of women were admitted at lower cervical dilatation. In both studies women admitted at lower cervical dilatation spent more time in the labour ward and received more labour interventions than women admitted at greater cervical dilatations.

Jackson et al. (2003) conducted a secondary analysis on data from a study which had compared two models of care; a collaborative obstetrician and certified nurse midwife model, where women gave birth in birth centres, and a traditional US private physician model, where women gave birth in large maternity units. This analysis aimed to compare the effects of model of care and timing of hospital admission on subsequent delivery outcome. Early admissions were defined as women who were admitted to hospital with a cervical dilatation of less than four centimetres and both primiparous and multiparous women were included. Almost 50% of the women in the traditional

model of care (private physician) were admitted early compared to 26% of women in the collaborative care model. For both models of care primiparous woman who were admitted early were significantly more likely to have assisted or operative delivery than those admitted later, with the highest level of assisted or operative delivery in the tradition model group who were also admitted early. This study suggests that early admission and model of care have an effect on labour outcome, however it was subject to a number of methodological flaws which make it difficult to draw conclusions about the effect of early admission. This was a retrospective study; although two models of care were compared there was no random allocation of women to group. Women choosing to give birth in a free standing birth centre were likely to have been quite different from women choosing traditional hospital based care. Although all the women in the study were of low income, there were significant demographic differences between the groups and this is likely to have affected both choice of birthplace and birth outcome. Women's expectations of childbirth and choice of place of birth have been shown to affect subsequent use of medical intervention and birth outcome (Machin and Scamell 1997; van der Hulst et al. 2004) and a number of RCT have demonstrated the effect of model of care (Hundley et al. 1994; Turnbull et al. 1996).

Klein et al. (2003) investigated whether the judgement policies of individual clinicians in relation to timing of labour admissions were associated with subsequent use of medical intervention and labour outcome. This study collected data retrospectively on 3485 primiparous, low risk women, and 133

family physicians responsible for care of normal healthy women in labour in one hospital setting. In this study doctors were classed as early admitters if they normally admitted at least 50% of women in their care who had a cervical dilatation of three centimetres or less, conversely late admitters admitted at least 50% of women at greater than three centimetres cervical dilatation. However, these groups were not particularly distinct, as during the study period 42% of the women admitted with a cervical dilatation of less than three centimetres were actually admitted by doctors classed as late admitters. The study found that women admitted by a doctor who was classed as an early admitter had higher rates of electronic fetal monitoring, epidural anaesthesia, caesarean or forceps delivery than women cared for by late admitters. Maternal factors, age, ethnic origin and, in particular, malposition of the fetus were also predictors of increased medical intervention in labour. The findings suggested that some doctors have a tendency to admit women to labour ward at lower cervical dilatation and that women cared for by these doctors are more likely to receive intervention in labour. However, the cervical dilatation of women on admission was not included in the data presented, so it is not clear how many women in each group were actually admitted early.

Rahnama et al. (2006) conducted a prospective study which examined the impact of early admission on method of delivery. The study included 810 primiparous women in spontaneous labour. Of these, 474 were reported to have been admitted during the latent phase of labour (not defined) and 336 during the active phase (presence of regular painful contractions and cervical

dilatation greater than three centimetres). Subsequent management of both groups of women was the same as follows; the woman was assessed two hours after admission, if her cervix had not dilated during that time artificial rupture of the membranes was performed and the woman was reassessed one hour later. If no cervical dilatation had occurred oxytocin was commenced. Women admitted during the latent phase of labour were significantly more likely to have a caesarean section than women admitted during the active phase. The main reason for operative delivery in women admitted during the latent phase was dystocia (slow progress of labour) and the median cervical dilatation for these women at caesarean section, was two centimetres or less. Unsurprisingly, Rahnama et al. (2006) conclude that most of these women were misdiagnosed as having labour dystocia when they were in fact still in the latent phase of labour. The authors go on to highlight the importance of accurate diagnosis of active labour and of admitting to the labour ward only women who are in active labour. This study is ethically and logically flawed in that the protocol required women diagnosed as being in the latent phase of labour to be managed as if in the active phase, only to report that 65.5% subsequently required caesarean section.

This group of studies comments on early admission of women and subsequent outcome although they are of mixed quality. Hemminki and Simukka (1986), Holmes et al. (2001) and Jackson et al. (2003) all describe clinical outcomes for women admitted early, as defined by cervical dilatation. These studies all used retrospective data analysis, and the samples of women and subsequent

outcomes would have been subject to a number of confounding factors. They suggest that women admitted early are more likely to receive intervention in labour, however causation cannot be determined. Nevertheless these studies may be hypothesis generating. Two main explanations are proposed: there could be factors intrinsic to some women, psychosocial or physical, such as pain, fear or lack of support, which lead them to seek early admission or factors intrinsic to the hospital admission itself which leads both to the early admission and to the higher levels of intervention. Suggested factors are misdiagnosis of active labour, physician preferences or that the clinician's assessment of labour progress is influenced by the amount of time the woman spends in the labour ward.

The notion of the influence of time on clinicians' judgement is supported by a study of doctor's decision-making (Cartmill and Thornton 1992), which found that merely altering the appearance of the duration of labour, while not changing the actual clinical information provided, encouraged doctors to intervene. Following labour ward admission, progress of labour is usually monitored graphically by plotting cervical dilatation over time (Friedman's curve); with a rate of one centimetre (or possibly 0.5 centimetres (NICE 2007)) per hour considered to be normal during active labour (Friedman 1989). Where this anticipated progress does not take place a diagnosis of labour dystocia or slow progress of labour is likely to be made; the main treatment of dystocia is amniotomy and oxytocin to augment labour progress (NICE 2007). Cartmill and Thornton (1992) used vignettes in which the graphical presentation of the

same clinical information was varied to produce the appearance of a longer labour, either by altering the ratio of time to cervical dilatation (flattening the curve) or by including or excluding the latent phase of labour. The results demonstrated that where the graphical presentation gave the appearance of a longer labour, or where the latent phase of labour was included, doctors were more likely to say they would intervene, either by recommending forceps or caesarean section delivery or by augmenting the labour with oxytocin. It has been suggested (Simonda 2002: Kitzinger 2006) that when a woman is admitted to the labour ward the clock starts ticking. Simonda (2002) proposes that rigid adherence to time provides a highly reliable organisational and cognitive order in which the hospital staff's need for predictability is satisfied.

Although it is not yet clear why women are admitted to labour wards while they are not yet in active labour, there is evidence that this affects a considerable number of women. Ball and Washbrook (1996) reported that up to 30% of admissions to UK labour wards were of women who subsequently turned out not to have been in labour. Klein et al. (2003) in the background to their study similarly reported that 30% of women were admitted to hospital with a cervical dilatation of two centimetres or less. More recently, Spiby et al. (2006a) surveyed maternity units in England and found the estimated rates of these admissions ranged between 10 and 100% (a finding which is hard to understand). In some of the units surveyed data were based on estimates and therefore may be unreliable, nevertheless they do suggest a considerable number of women are admitted while not yet in active labour.

2.4.3. Early admission and labour intervention – summary

Across the developed world the rate of medical intervention in labour has risen and is an issue of widespread concern. There is evidence of a cascade effect of intervention and that women who are admitted to labour wards early are more likely to receive intervention in labour than those who are admitted later. It has been suggested that a possible reason for this higher rate of intervention is that clinicians do not make an accurate distinction between women who are in active labour and those who are not yet in labour, or who are in the latent phase. This may be because they misdiagnose active labour or because they use labour ward admission itself as a proxy measure for active labour. Once admitted, the mere presence of a woman in the labour ward over a protracted period of time may encourage caregivers to intervene.

Concern over the high levels of intervention experienced by these women, as well as the economic cost associated with inappropriate admissions, has led to the development of a range of systems which aim to improve the management of early labour, in particular to reduce unnecessary labour ward admissions (Spiby et al. 2006a). The next section of the literature review provides an overview of three such systems; triage, home assessment and clinical pathways and guidelines.

2.5. Systems of early labour management

One of the most commonly implemented admission management systems is triage (Angelini 2000) which is a means of prioritising patients in order of their care needs. Triage was originally developed for battlefield medical settings where it was used to prioritise casualties in order to maximise survival (as opposed to merely treating the most seriously injured first) (Mahlmeister and van Mullem 2000). It was initially introduced to civilian healthcare systems in the United States to prioritise treatment in emergency departments (Berman et al. 1989; Brillman et al. 1996). The purpose of triage is not specifically to diagnose or to treat but rather to assess patient need and make appropriate referral (Berman et al. 1989). Much of the literature on triage in maternity care relates to the North American medical system, this is unsurprising as obstetric triage is now a legal requirement for hospitals that participate in the Medicare programme (almost all US hospitals) (Mahlmeister and van Mullem 2000). Triage in maternity care has been described as an efficient patient care delivery system useful in a high volume obstetric unit (Zocco et al. 2007). Early labour assessment is reported to be one of the most common, although not the only reason for the use of triage in maternity care (Austin and Calderon 1999; Kelly 1999; Angelini 2000; Spiby et al. 2006a). Telephone triage has also been introduced as a means of screening women to identify those who require face-to-face consultation (Spiby et al. 2006a).

Despite the universal use of triage in the United States and the increasing use of triage elsewhere, there has been little evaluation of its efficacy in maternity

care. Spiby et al. (2006a) conducted a survey of maternity units in England and reported that almost 9% had introduced telephone triage while 10% had a designated triage facility. This survey described the experience of triage services in England and concluded there was no definitive evidence of their effectiveness. Spiby et al. (2006a) also conducted a qualitative evaluation of the telephone triage component of the All Wales Clinical Pathway for Normal Labour (NHS Wales 2006) reporting that midwives were generally accepting of the pathway, viewing it primarily as a means of standardising current practise, and that women's experiences while generally positive were more variable with some women reporting dissatisfaction with the service (Spiby et al. 2006a).

Studies of triage in other clinical settings have found considerable variation in levels of consistency in the triage judgements of clinicians (Brillman et al. 1996; Considine et al. 2000), lack of correlation between triage decisions and the experience of practitioners, and that nurses use heuristics (rules of thumb) and intuition in making triage decisions (Cioffi 1998). Cioffi (1998) suggests that practitioners in triage settings make decisions under conditions of uncertainty. Triage is a patient management system that provides a setting in which clinical judgement may take place. While good clinical judgement is essential in triage (as in any clinical area) the use of triage in itself does not specifically support judgement.

2.6. Home assessment and support

Early labour assessment at home has not been widely implemented in the UK. It did form part of the DOMINO (Domiciliary In and Out) system of care which was widely available in the UK through the 1980s and 1990s (Murphy –Black 1992; Wardle et al. 1997). Although uptake of DOMINO care was generally low (McClean et al. 1999; Wardle et al. 1997). A similar model of care is currently used in some rural areas of Scotland and home assessment is used by independent midwifery practices. Until recently there has been little evaluation of home assessment schemes as part of normal maternity provision.

A randomised controlled trial (RCT), conducted in Canada, compared telephone triage with home assessment of labour (Janssen et al. 2003). Women in the home assessment group received a physical assessment of labour which included specific diagnostic criteria as well as emotional support and advice, while those in the telephone triage group received telephone advice alone. Fewer women in the home assessment group were admitted in the latent phase of labour. They also required less narcotic analgesia and were more satisfied with their care than those in the telephone triage group. However, there was no difference in the use of oxytocin for augmentation of labour or other labour interventions. The findings of this study suggest that telephone triage is less effective in reducing early admissions than home assessment and that a face-to-face encounter which may include a physical examination, is preferable both in terms of clinical outcomes and women's satisfaction.

The Early Labour Support and Assessment trial (ELSA) currently underway (Spiby et al. 2006b) is a randomised controlled trial which aims to investigate the impact of providing midwifery support to primiparous women in early labour in their own home compared to standard care. This trial is expected to report in 2008.

Home assessment does appear to be a more appealing option than triage in that it has the potential to provide an individual consultation in the woman's own home rather than the process-production model offered by triage. Home assessment, like triage, is a system of care in which clinical judgement is required, although the need for a management decision (remain in hospital or discharge home) may be deferred and this may allow time for a 'wait and see' labour diagnosis. However, home assessment is likely to be resource intensive and may not prove to be cost effective if offered as part of standard maternity service provision.

2.7. Clinical pathways and guidelines

Clinical pathways, guidelines and protocols have proliferated in healthcare over the last ten to fifteen years. Terminology has been used interchangeably. However, their overarching purpose is to provide evidence based guidance for care in specific clinical situations. For example, clinical pathways have been described as structured multidisciplinary plans of care designed to support clinical management and encourage translation of evidence based guidelines into clinical practice (Campbell et al. 1998; Hunter 2007). Guidelines and

pathways may contain explicit decision support or may provide more general guidance (decision support tools are discussed in chapter three). Several recent UK guidelines have addressed care of a woman in normal labour (RCM 2005; NHS Wales 2006; NICE 2007). The RCM guideline provides general evidence based advice about care in labour without containing specific criteria for diagnosis of labour. The All Wales Clinical Pathway for Normal Labour (NHS Wales 2006) and the National Collaborating Centre for Women's and Children's Health intrapartum care guidelines and care pathway (NICE 2007) both contain specific criteria for diagnosis of active labour. Both of these documents acknowledge that there is no firm evidence for the definition of established or active labour recommended ('cervical dilatation of more than three centimetres and fully effaced in the presence of regular painful contractions' (NHS Wales 2006 p4) and 'regular painful contractions and progressive cervical dilatation from four centimetres' (NICE 2007 p138)).

A criticism of guidelines and pathways is that they are rarely rigorously evaluated (Campbell et al, 1998; Hunter 2007). Further, because they tend to contain a number of elements of guidance, it may then be difficult to determine which aspects of the guideline or pathway has been effective (i.e. specifically what the active ingredient is). For example, the evaluation of the All Wales Pathway (Hunter 2007) concluded that the pathway was a 'complex multifaceted intervention' which had 'complex and unexpected outcomes on the experiences of midwives, mothers and doctors' (p2) while apparently having little impact on clinical outcomes (Hunter 2007).

2.8. Systems of early labour management – summary

Each of the systems described have been implemented to address the problem of inappropriate admission of women to labour wards and to improve the management of early labour on the premise that this will, in turn, reduce the use of unnecessary intervention in labour. These systems may provide a setting in which the diagnostic judgement will take place (triage or home assessment) or a vehicle for decision support (guidelines and pathways) however they do not in themselves, ensure good clinical judgement. It follows that whether early labour assessment is conducted in the home, labour suite or triage area, good diagnostic judgement by clinicians is essential (in particular, where hospital birth is the predominant model) so that women are not admitted to labour wards before they are in active labour. However, there is considerable evidence that diagnosis of active labour is not a simple judgement and that misjudgements are frequently made (Hemminki and Simukka 1986; Ball and Washbrook 1996; Holmes et al. 2001). The final section of the literature review explores possible reasons for its difficulty.

2.9. Labour and uncertainty

It is unsurprising that diagnostic misjudgements may be made as there is uncertainty about several fundamental aspects of labour onset, in particular, when normal labour will start, why labour starts and what the parameters of normal labour are.

Pregnant women are usually given an expected date of delivery during their antenatal care. However, accurately predicting the likely timing of normal labour onset is still uncertain (McLean 2001; Smith 2001; Hollis 2002). Gestation of pregnancy is anticipated to be between 259 and 293 days and normal labour is expected to start at term, a five week period between the end of the 37th and end of the 42nd week of gestation (WHO 2007). Until the middle of last century, when obstetric ultrasound was developed, accurate confirmation of the time of conception was difficult and was the subject of legal as well as medical debate (Ballantyne and Browne 1922; Kortenoever 1950; Stewart 1952). While the use of ultrasound now permits dating of the start of pregnancy, predicting when labour will start has been confounded by the practice of offering routine induction of labour in pregnancies which continue beyond 41 weeks (Gülmezoglu 2006). This has meant that it is rare for a pregnancy to continue into the 43rd week of gestation and therefore the average duration of pregnancy cannot be accurately assessed (Smith 2001). Thus the expected date of delivery is more accurately described as an estimated date, a detail which may lead to uncertainty for clinicians and mothers alike.

The mechanisms that trigger the onset of labour are also poorly understood (Greer 1995). There are three main theories; the progesterone block theory, first suggested by Csapo (1961); the maturation of the fetal hypothalamic pituitary adrenal (HPA) axis theory (Liggins et al. 1967 cited by McLean 2001; McDonald and Nathanielsz 1991 cited by Nathanielsz 1994) and the 'placental clock' theory which suggests that the maturation of the fetal hypothalamic

pituitary adrenal (HPA) axis is initiated by the release of placental hormones. An overview of these theories is presented in Appendix 1.

While the factors which initiate labour are uncertain, the physiology of labour onset is well understood (Bishop 1964; Anderson and Turnbull 1969; Caldeyro Barcia 1959 cited by Greer 1995; Uldbjerg et al. 1983; Ulmsten 1997). The onset of labour is not sudden but a gradual process which takes place over the last few weeks of pregnancy, a period termed pre-labour (Greer 1995). In the cervix a process described as cervical ripening takes place in which the cervix changes in consistency from a firmly closed ridged ring to a softer tissue capable of dilatation by direct muscle traction and the force of the presenting part during labour (Greer 1995; Ulmsten 1997). At the same time there is a change in the myometrium from the largely relaxed state which accommodates the stretch required during pregnancy to one capable of the co-ordinated contraction and retraction of muscle characteristic of labour (Nathanielsz 1994). Intracellular connections known as Gap Junctions form (Garfield et al. 1977; 1980; Greer 1995; Challis et al. 2000; Keelan et al. 1997; Nathanielsz 1994) and these allow the uterus to contract in a co-ordinated manner.

Labour itself is a process, a transition which is primarily, although not exclusively physical (Greer 1995; Wagner 1998; Downe 2001) and is described as having three main stages. While the end of the final stages of labour are quite easily defined (the birth of the baby marks the end of the second stage

and the expulsion of placenta and membranes the end of the third), there is much less certainty about the parameters of the first stage of labour.

There is broad agreement that the first stage of labour, which ends with full dilatation of the cervix, may be further divided into two phases, the latent and active phases (Austin and Calderon 1999). The latent phase is a poorly defined period from the onset of regular contractions to the start of the active phase, the phase of increased rate of cervical dilatation. However, the duration of the latent phase is difficult to predict (Crowther et al. 1991), contractions of the latent phase may be confused with Braxton-Hicks contractions (non-rhythmic, pre-labour contractions) and an increased rate of cervical dilatation may only be determined retrospectively.

The WHO defines the signs of labour onset as; painful regular contractions, effacement and or dilatation of the cervix, rupture of membranes, bloody discharge (WHO 1997). While emphasising the central importance of accurate diagnosis of labour the WHO guidance is somewhat vague, for example signs are not prioritised and it is not clear whether all signs are required for diagnosis of active labour or whether one key sign would be sufficient. Diagnosis of active labour was a key component of the active management of labour package, introduced in Dublin in the 1970s (O'Driscoll et al. 1973; O'Driscoll and Meagher 1980), and strict diagnostic criteria or cues were presented. These were; regular painful contractions as well as one of the following; bloody discharge or show, spontaneous rupture of membranes (SRM) or dilatation of

the cervix to at least three centimetres. Most subsequent studies which have included diagnosis of active labour have referred to this work and have incorporated the same or similar criteria describing the cues as part of an evaluation of the larger active management of labour package or components of that package (O'Driscoll et al. 1984; Turner et al. 1986; Boylan et al. 1991; Lopez-Zeno et al. 1992; Frigoletto et al. 1995; McNiven et al. 1998; Sadler et al. 2000; Lauzon and Hodnett 2001; Janssen et al. 2003).

Considering the specific cues described, all of the studies identified the presence of painful uterine contractions as the primary cue for diagnosis of labour, all included dilatation or effacement of the cervix; three gave equal weighting to the presence of show (O'Driscoll et al. 1984; Turner et al. 1986; Frigoletto et al. 1995) and five spontaneous rupture of membranes (O'Driscoll et al. 1984; Turner et al. 1986; Lopez-Zeno et al. 1992; Frigoletto et al. 1995; Sadler et al. 2000).

Only one study has specifically evaluated the use of explicit criteria for active labour diagnosis (McNiven et al. 1998), this study was the only inclusion in a Cochrane Review of labour assessment programmes to delay admissions to labour wards (Lauzon and Hodnett 2000; 2001). In this study 209 low risk women were randomly allocated when they presented in spontaneous labour. All women in the control group were admitted directly to the labour ward without prior labour assessment, while women in the experimental group had their labour assessed based on the presence of regular painful contractions and cervical dilatation greater than three centimetres. Women judged not to be in

labour were sent home or remained in an assessment area to await the establishment of active labour. The study reported that when labour was assessed using strict diagnostic criteria significantly fewer women received oxytocin to augment labour and less pain relief was used compared to no labour assessment (22.9% compared to 40.4% and 7.6% compared to 20% respectively). These results suggest that using specific criteria to diagnose active labour may reduce the number of women who are admitted to the labour ward while not in labour or in the latent phase of labour and that this may result in reduced intervention in labour. However the study had a sample size of only 209 women, conducted in one hospital, therefore the results cannot be used to draw definite conclusions.

A descriptive study focussed specifically on midwives' experiences of diagnosis of labour and cues for subsequent labour progress (Burvill 2002). In this study a focus group was conducted with midwives and this was followed by in-depth interviews with one experienced midwife. Burvill (2002) proposed a model of midwifery diagnosis of labour which acknowledged the subtle changes which take place toward the onset of labour incorporating cues from late pregnancy such as 'nesting activity', 'excitement' and 'bright shining eyes', through to active labour in relation to women's reactions (e.g. mood, energy and movement), external signs (e.g. breathing, contractions, show and appearance of a red line between the buttocks) and internal signs (e.g. cervical dilatation and condition of membranes). This study is interesting in that it identifies the non-clinical features which midwives may use when assessing a woman.

However the findings are based largely on cues identified by one midwife which have not been tested for efficacy and would clearly depend on a close relationship being established between the woman and her midwife for some time prior to labour onset.

2.10. Labour and uncertainty – summary

Making a distinction between a woman who is in active labour and one who is not is an important clinical judgement, in particular where hospital birth is the predominant model. There is evidence that significant numbers of women are admitted to labour wards while not yet in active labour and that misdiagnosis of active labour contributes to these admissions. However, labour itself is a process; the start of the active phase is merely a point in this process. In addition, there are a number of aspects of the onset of labour which are not fully understood, these factors increase the level of uncertainty in the judgement situation.

The cognitive continuum theory (Hamm 1988; Hammond 1996) proposed a model of judgment in uncertainty in which a person uses available cues in order to make inferences about the situation which cannot be seen (the theory is described in chapter three). Some cues will have a high degree of correspondence (or salience) with the situation to be inferred, while others will have weak correspondence. It appears that diagnosis of active labour is this type of judgement, in which the clinician makes a prediction about future progress of labour based on the assessment of current informational cues.

Overall, the literature reviewed indicated consistency in regard to the cues described for diagnosis of labour, with most of the studies identifying the same or similar physical cues (regular, painful uterine contractions, cervical dilatation and effacement, ruptured membranes and bloody show) although with a range of levels and combinations. These cues all have high correspondence with the physiological changes which occur at labour onset as described above (for example, cervical ripening and changes in myometrial contractility). However, they are not directly diagnostic in the way, for example, that low haemoglobin is diagnostic of anaemia. The cues presented by Burvill (2002) such as changes in breathing and movement do relate to labour onset but they have a weaker degree of correspondence. There are a number of possible situations, in addition to labour, in which a pregnant woman could be breathless or have restricted movement. In this type of judgement situation the cognitive continuum theory suggests that people use intuitive judgement styles which may be prone to error and that by changing the judgement task, for example by introducing decision support, the accuracy of the judgement may be improved.

In the next chapter an overview of judgement and decision-making literature is presented and rational and non-rational theories of judgement are discussed. The evidence for the contribution of decision support to clinical judgement is described and this forms the theoretical background for the choice of an algorithm as the intervention in this research.

CHAPTER 3: AN OVERVIEW OF JUDGEMENT AND DECISION-MAKING THEORIES

Deciding whether active labour has commenced is acknowledged to be one of the most difficult aspects of the care of a woman in labour (O'Driscoll et al. 1973). There is considerable evidence, described in chapter two, of the important clinical and resource implications of misdiagnosis (Hemminki and Simukka 1986; Holmes et al. 2001; Klein et al. 2003; Jackson et al. 2003; Rahnama et al. 2006), as well as evidence suggesting that many women are admitted to labour wards in the UK, and other developed countries, while not yet in established labour (Ball and Washbrook 1996; Spiby et al. 2006a). Midwives are key care givers in the provision of maternity care to normal healthy women in the UK, and in many countries. However, despite the importance of the judgements and decisions that they make, there has been little research on the way in which midwives make judgements or on how they may be optimised. This chapter provides an overview of the main theories of judgement and decision-making and describes how judgement and decision-making theory has contributed to the development of this research. The focus is in particular, on individual practitioner judgement. Theories of shared and group decision-making are not discussed.

3.1. Judgements and decisions

A judgement is described as the assessment of alternatives and a decision as a choice between alternatives (Tversky and Kahneman 1974; Dowie 1993). The process of identifying whether or not a woman is in active labour may be

described as a diagnostic judgement (Swets 2000) as the clinician makes a prediction about future progress of labour, or alternatives such as Braxton Hicks contractions or urinary tract infection, based on the assessment of current informational cues. The subsequent management, for example, whether the woman should be admitted to a labour ward or discharged home, would be considered to be a decision.

3.2. Research in judgement and decision-making

Judgements and decisions are ubiquitous aspects of every day life. Hammond (1996) describes the importance of judgement as follows:

> Human judgement is the hidden, mysterious link in the process that forms the policies and plans that directly effect, if not control the nature of our society, as well as its interaction with other societies. (Hammond, 1996 p 5)

Underlying the need for judgement is the notion of uncertainty. Most judgements made in everyday life, are made based on information that is unclear, 'noisy' and uncertain. Judgements frequently must be made under time pressures and in an emotional atmosphere in situations where there may be organisational and resource limitations as well as unpredictability of outcomes. This has been described as judgement under conditions of irreducible uncertainty (Dalgleish and de Michele 1995; Hammond 1996; Hastie

and Dawes 2001). Considering a situation in which it was possible to gather all necessary information to make an accurate prediction of future events then outcomes for actions would be known and the exercise of judgement would not be necessary (Hacking 1990). However, for most judgement tasks it is not practical, or possible, for all necessary information to be gathered at the time that the judgement must be made, therefore uncertainty cannot be completely removed and the application of judgement is essential (Hammond 1996).

Good judgement and decision-making is central to the provision of high quality health care, and is a key aspect of health policy. For example, judgements about risk in pregnancy and decisions about antenatal screening or type of delivery are considered essential in maternity service provision. However, there is little guidance for midwives or obstetricians on what constitutes good judgement or how decisions may be improved. The field of judgement and decision-making research has addressed these, and similar, issues across a wide range of subject domains by considering three main research approaches; descriptive, normative and prescriptive. Descriptive approaches attempt to explain how people make judgements and decisions, normative techniques describe how people should, ideally, make judgements and decisions while prescriptive techniques aim to improve the judgements and decisions that people make (Baron 2000).

3.3. Analysis and intuition

Pre-dating modern research in judgement and decision-making, two fundamentally distinct types of thinking were recognised (Hammond 1996; Kahneman and Frederick 2002). Although many different terms have been used to describe these, they may be broadly classified as analysis and intuition. Analytical thinking is characterised as slow, reasoned and deliberate thinking which may be logically explained, while intuitive thought is fast, automatic, experiential and may not be logically described or explained. Often described as mutually exclusive, the distinction between these two modes of thinking has a parallel in judgement and decision-making theory where a fundamental distinction has been made between rational judgement models and intuitive or non-rational judgement models (Hammond 1996; Gilovich and Griffin 2002). This distinction may also be found in nursing and midwifery literature where some authors differentiate between alternative 'ways of knowing', rational or scientific knowledge, and intuitive or craft knowledge (Davis– Floyd and Davis 1997; Paley et al. 2007).

3.4. Rational judgement

In judgement and decision-making literature rational models propose that judgement choices are made based on the principles of probability (that is, a mathematical approach to events characterised by randomness or uncertainty), and utility (the degree of worth attached by an individual to a particular outcome), (Hastie and Dawes 2001). Using rational judgement a person would assess the probability and utility of each option and make a judgement based

on their optimal combination thereby maximising personal utility (Gillovich and Griffin 2002).

Expected utility theory, first proposed by von Neuman and Morganstern (1947) (cited by Hastie and Dawes 2001) as a normative theory, forms the basis of the most frequently used models of rational judgement. The key principles of expected utility theory are:

- A rational person will make choices between alternatives by following rational rules.
- An ordering of choices may be made i.e. the decision maker will prefer one option over another, or be indifferent.
- It is possible to assign a numerical value to each possible outcome or consequence; this will be the utility of each possible consequence.
- The expected utility is then calculated from the sum of possible utilities and the probability of occurrence of each.
- The option chosen will be the one with the highest expected utility (Hastie and Dawes 2001).

Bayes Theorem, a form of mathematical probability, was introduced by Edwards et al. (1963) as a means of revising the probability of an event given the evidence acquired i.e. as new information is gained. Much research in judgement and decision-making has been concerned with identifying and explaining the ways in which normal human judgement either conforms or

departs from the rational principles proposed by this ideal model (Gillovich and Griffen 2002).

A development of expected utility theory was subjective expected utility theory (Edwards 1954), which acknowledged that normal human judgement may fall short of rigorous rationality, as described above, and that probabilities may require to be subjectively estimated. For example, a paediatrician may estimate a baby's chance of surviving if born at 28 weeks as about 85% based on personal clinical experience, rather than empirical evidence of survival rates. Although these theories are primarily normative, some authors suggest that humans are inherently rational in their judgement and decision-making (Edwards 1968, cited by Gilovich and Griffin 2002), that the mind works in essentially the same way as Bayes Theorem, and that people form judgements in everyday life based on rational principles.

The theory of bounded rationality (Simon 1957 cited by Gilovich and Griffin 2002) sought to explain discrepancies between optimal rational judgement and observed human judgement. It acknowledged that full rationality is an unrealistic descriptor of normal human judgement. The theory suggested that while people may think in a rational manner they must work within the limits of their cognitive abilities. Bounded rationality proposed that as people are limited in their capacity for computation they will use approximation, determining an option which is satisfactory based on one salient cue, rather than performing the mathematical calculations required by probability theory in determining the

optimal choice. In a vignette-based study of general practitioner (GP) decisions to prescribe lipid–lowering drugs for a set of hypothetical patients, Dhami and Harries (2001) identified that GPs adopted a simplifying strategy, using one key cue as a personal decision rule. For example, 'Does the patient have a cholesterol level of between 7.6 and 8.0? If yes, then prescribe the drug; otherwise do not prescribe'.

The concept that rational choice and maximised personal utility is a descriptor of real life judgements and decisions is, in particular, used by researchers in economics. For example the standard gamble technique, commonly used in health economics evaluation is based on expected utility theory (Ryan et al. 2001). Cairns et al. (1996) used a series of standard gamble questions as a means of exploring decision-making regarding antenatal screening. The study aim was to identify the utility individual women would place on two possible options for antenatal screening, or no screening. The study found that for the majority of participants, expected utility would be maximised by accepting screening (although it was unable to identify a preference for type of screening), therefore, if individual women were making decisions following the principles of rationality, they would accept the antenatal screening.

Information on expected utilities (Cairns et al. 1996) could, theoretically, be incorporated into prescriptive decision-making techniques (methods which aim to improve judgement and decision-making). Models such as subjective expected utility theory (Edwards 1954), and multi-attribute utility theory

(Edwards and Newman 2000) have been developed as prescriptive models, often presented as complex decision trees in which the probability and utility of each potential choice option may be mapped (Dowie 1996; Dowding and Thompson 2002). It has been suggested that these models may improve judgements and decisions in a range of real world settings (Letourneau and Jensen 1998). However, they require a high level of numeracy, an understanding of probability and a considerable amount of time, both for development and implementation. They are potentially applicable in improving planned and shared decisions where time is available to calculate probabilities and elicit the personal utilities of stakeholders. For example, Dowding and Thompson (2002) developed a decision tree as an exemplar of a woman's decision whether or not to undertake antenatal screening. However, these techniques are less likely to be useful in situations where the time available to make the judgement or decision is limited.

Although it is clear that humans are capable of making rational judgements when they have the knowledge and the tools, detractors of rational choice theories argue that it is not the normal means of human judgement (Kahneman et al. 1982). They suggest that it ignores the limits of humans for computation, suggesting that subjective expected utility theory is not a good descriptive or normative theory for decision-making (Frisch and Clemen 1994) and that it is impossible to apply in making actual decisions (Simon 1983). Further, there is a lack of empirical evidence on which to calculate probabilities for many clinical situations and even where these are available, research has shown that people

are not good at understanding them (Gigerenzer 2002). In a recent study which investigated the accuracy of interpretation of probabilistic screening information, participants (obstetricians, midwives, pregnant women and their partners) were asked to estimate the probability that a baby had a genetic condition, given a positive antenatal screening test. Participants were given information about base rate of the condition and positive and negative predictive value of the test, either in the form of percentages (e.g. 1% of babies have Down's syndrome) or frequencies (e.g. 100 out of 10,000 babies have Down's syndrome). The study found that overall, most responses (86%) were wrong, across all groups, and that while responses for obstetricians were more accurate using frequencies, only 34% were correct. Midwives' responses were the least accurate; none of the midwife participants gave correct answers for either form of data presentation (Bramwell et al. 2006).

In addition, personal utilities may vary widely between individuals and are difficult to ascertain. For example, an obstetrician may believe that a caesarean section is the optimum delivery choice for a woman where dystocia has been diagnosed, while the woman believes a normal delivery is the only acceptable option. In an alternative situation a midwife may feel a woman should strive for a normal birth, while the woman wishes to have a caesarean section which she believes will protect her pelvic floor. Computerised decision support may offer the possibility of rapid processing of probabilities and even of including multiple stakeholders' utilities (Dowding and Thompson 2002). However, it seems likely that these systems will only ever be produced to

support a few high level clinical decisions (for example, whether to undertake elective surgery) rather than the numerous small decisions which make up the bulk of midwifery practice (for example, whether or not to perform artificial rupture of membranes, whether to breast or bottle feed).

3.5. Non-rational models

The heuristics and biases approach to judgement, first proposed in the 1970s (Kahneman et al. 1982), was a radical departure from rational choice theory. This descriptive theory proposed that judgements made in every day life do not conform to the laws of probability, but rather are based on a set of simplifying heuristics – or rules of thumb. This, it is suggested, is not an error of rational choice, but is a fundamentally different cognitive process, which is an efficient and fast means of making judgements. However, these types of heuristic-led judgements are prone to systematic biases which lead to error.

Three principal heuristics were originally described; Representativeness, Availability, and Anchoring and Adjustment (Kahneman et al. 1982). Although other heuristics have subsequently been suggested these three remain the foundation of the heuristics and biases approach. Cioffi and Markham (1997) in a study of midwifery judgement and decision-making suggested that midwives use heuristics such as representativeness and availability when making judgements about women admitted to their care. Each heuristic has a corresponding set of biases which describe the way in which judgements may depart from the normative standard of the rational choice model and the laws of

probability. These heuristics and examples of their related biases will be described in turn.

3.5.1. Representativeness

Using the representativeness heuristic, people are said to match the characteristics of an object with their stored mental models or prototypes. This means that their judgement of the likelihood of an object belonging to a category will depend on the similarity between that object and the stereotypical characteristics of the category. For example, a person will be judged likely to belong to a particular group if they resemble the stereotypical members of that group. A number of biases are associated with this heuristic, including base rate neglect, belief in the law of small numbers and misconceptions of chance. Base rate neglect, describes the bias which is created when people do not take into consideration the effect of base rate in estimating probability. For example, a newly qualified midwife working in a small maternity unit in Scotland with 400 annual births may fear the occurrence of severe post partum haemorrhage despite the risk of such an event being less than 10 per 1000 births (SPCERH 2001). Belief in the law of small numbers is characterised as judgements which do not take into account the effect of small sample size, leading to overconfidence in the outcomes of small samples when generalising to the corresponding population. Misconceptions of chance, describes the judgement error which occurs because people tend to expect that a short randomly generated sequence will have the same characteristics as a large randomly generated sequence. So, for example, in tossing a coin six times, it might be

expected that the sequence HTHHTT might be more likely than the sequence HHHHHT.

3.5.2. Availability

Using the availability heuristic the likelihood of an event is predicted by the ease with which similar cases may be recalled. In this situation common events are usually more easily retrieved from the memory than rare events (hear hooves, think horses not zebras). However, the biases which have been associated with this heuristic indicate that it is not invariably reliable (Hastie and Dawes 2001). For example, a situation or event which is easily recalled will appear more common than an event which is more mundane and less easily recalled, this bias is known as retrievability of instances (Hastie and Dawes 2001). Thus dramatic or exciting occurrences may be judged more likely to occur than they do. For example, people may over estimate the likelihood of fatal shark attack because of the publicity surrounding such instances. Conversely, the likelihood of death in household accidents may be underestimated because of the more mundane nature of these common incidents (Hastie and Dawes 2001). Life or work experience dictates that some events will be experienced in our own setting more commonly than in the general population and therefore may appear more prevalent, this bias is known as structural availability. For example, a nurse or midwife working in a neonatal unit may believe that the incidence of genetic abnormalities is high because many of these babies will require admission to neonatal units.

3.5.3. Anchoring and adjustment

The anchoring and adjustment heuristic suggests that in making judgements, people tend to take an estimated starting point and adjust their estimate either up or down to come to a final answer. In some situations of uncertainty an inaccurate starting point may be selected, nevertheless, this forms an anchor and subsequent adjustments are usually insufficient, remaining biased toward the anchor point (Tversky and Kahneman 2000). The status quo is one of the most powerful anchor points in every day life (Hastie and Dawes 2001). Tversky and Kahneman (2000) suggest that even when an arbitrary anchor point is suggested subsequent adjustment will be insufficient demonstrating bias toward the initial point. In addition, they suggest that once an initial start point, or hypothesis, has been established this biases subsequent information seeking, for example, a doctor who first suspects a preliminary diagnosis may only conduct tests which confirm that diagnosis. This is known as confirmatory bias.

The heuristics and biases approach has been very influential, however it has been criticised. In particular, it has been suggested that it presents an overly pessimistic view of human judgement (Ortman and Hertwig 2000, cited by Gilovich and Griffin 2002), while in reality there is considerable evidence of the success of human judgement. A second commonly cited criticism is that the heuristics and biases research programme is frivolous and that the experiments have been conducted in such a way as to manipulate subjects to obtain evidence of systematic errors in reasoning (Gigerenzer 1991). Tversky and

Kahneman (1983) defend their model, arguing that heuristics should be viewed as natural assessments rather than deliberate, and lazy, mental short cuts.

Despite its critics, the heuristics and biases approach is appealing as a model of judgement which has a high level of descriptive validity. However, it is primarily a descriptive theory, it does not offer a means by which judgement may be improved, other than the rather wishful notion that an understanding of the heuristics used may allow the decision maker to reduce their reliance on, or avoid, the corresponding biases (Hastie and Dawes 2001).

Both the rational and non-rational models of judgement outlined above provide an either or approach to human judgement. However, theories have also been proposed which acknowledge that humans may be both intuitive and analytical thinkers. These are the cognitive continuum theory (Hamm 1988; Hammond 1996) and the dual processing theory (Sloman 1996; Kahneman and Frederick 2002). These theories suggest that rational and non-rational modes of cognition are not mutually exclusive and that humans are capable of using either type of thinking depending on the characteristics of the judgement task at hand.

3.5.4. Cognitive continuum theory

The cognitive continuum theory was developed by Hammond (1996) based on the earlier work of Brunswik (1952), who proposed a descriptive model of intuitive judgement (the Lens Model) in which a person uses available data

(cues) in order to make inferences or judgements about a situation which cannot be seen. Some cues will have a strong correspondence with the situation or event to be inferred, these cues are described as having a high degree of ecological (or real life) validity. Other cues will have a weak correspondence, these have low ecological validity. For example, for a primiparous, pregnant woman, a high diastolic blood pressure has a strong link with pre-eclampsia, and so this would be described as a cue with high ecological validity, while ankle oedema has only a weak link and therefore would have low ecological validity. Brunswik (1952) suggested that it is a lack of cognitive awareness of the way in which these cues are utilised and integrated which makes a judgement intuitive. He further suggested that analytic cognition may be 'mellowed' by intuition and that in this way intuition and analytic cognition have a moderating effect on each other. Thus, he suggested that human cognition is a mix of analysis and intuition rather than all of one or the other. Brunswik termed this type of thinking quasi-rationality.

The cognitive continuum theory (Hammond 1996) likewise suggests that intuition and analysis are not mutually exclusive forms of cognition, but rather are at the opposite ends of a continuum. Further, it suggests that most judgements contain some elements of analysis and intuition and recognises the central place of common sense or quasi-rationality as the most frequently used form of human judgement. Hammond defines a common sense judgement as one which contains as much analysis as possible and as much intuition as necessary (Hammond 1996). Quasi-rationality is the cognitive mode which sits

in the middle of the cognitive continuum between the extreme ends of intuition and analysis.

Central to this theory is the notion that, parallel to the cognitive continuum is a task continuum on which different types of judgement tasks may be ordered. Hammond (1996) suggests that no mode of thinking is essentially better than any other and that it is the nature of the judgement task which will dictate the cognitive mode employed. Characteristics of the judgement task which encourage intuitive thinking are: the presence of a large number of cues (>5), cues presented simultaneously rather than sequentially, absence of a decision rule and lack of time in which to make a judgement. For example, a midwife's judgement about whether a woman has dystocia in the second stage of labour would involve processing cues such as the strength and frequency of contractions, the effectiveness of maternal pushing, the type of pain relief being used, the decent and position of the fetal head, maternal wellbeing, the fetal heart rate, and the length of time that the woman had been in the second stage of labour. This judgement task would typically be carried out in a busy labour ward where time pressure is a feature and anxiety about consequences of wrong judgements are paramount. Conversely, task characteristics which encourage analytic thinking are: fewer cues (2-4) presented sequentially, high ecological validity of cues, an agreed decision rule which allows the cues to be organised in a consistent manner and increased time available for the judgement. An example of this type of judgement would be that made by a midwife when caring for a primigravid woman who presents at an antenatal

clinic at 28 weeks gestation with a diastolic blood pressure of 110 mmHg and significant proteinuria. Here the midwife has fewer cues and more time to consider the appropriate course of action. While the cognitive continuum as described by Hammond (1996) is a descriptive theory, Hamm (1988) suggested that there is an appropriate cognitive mode for every judgement situation, that it is the wrong choice of cognitive mode which leads to inaccuracy and that a clinician could use the cognitive continuum to improve judgement. While it may be difficult to consciously change modes of thinking, it may be possible to alter the features of the judgement task, thereby inducing more analytic cognitive modes. Hamm (1988) suggested that this could be done by increasing the time available, reducing the number of cues used, removing redundant cues and applying a decision rule which structures the judgement task.

3.5.5. Dual process theory

Similarly, dual process theory (Sloman 1996; Kahneman and Frederick 2002) identified two systems of reasoning which are commonly referred to as System One and System Two. System One is characterised by intuitive, fast, associative, unconscious judgements, while System Two is deliberate, controlled, rule governed and involves slow, conscious reasoning. Dual process theory suggested that these systems are continually active and that the role of either type of cognition depends on the features of the judgement task. A number of factors have been suggested to be influential, including time available (Finucane et al. 2000), mood (Bless et al. 1996), and intelligence (Stanovich and West 2002). For example, the need to reduce cognitive load may dictate that System One be used in situations where a number of mental

tasks are required simultaneously since System Two requires greater mental effort and is only capable of handling one problem at a time (Kahneman and Frederick 2005). For example a nurse or midwife who is engaged in a drug administration round may give an intuitive answer to a question from a colleague on a different topic.

Dual processing and cognitive continuum theories are similar in that both suggest that judgement may encompass analytic and intuitive thinking. However, while cognitive continuum theory proposes that neither mode of thinking is superior and that a mix is the most common form of thinking. Dual process theory suggested that the two systems are neurologically and operationally distinct and that System Two has a supervisory or corrective role (Tversky and Kahneman 1971). More recently, Paley et al. (2007) has argued that there is little evidence to support the equal partnership relationship between System One and System Two. However, dual process theory is currently presented only as a descriptive model which does not suggest a method by which analytic judgement styles (System Two) may be induced.

3.6. Clinical judgement

A distinct theoretical approach has been used by researchers who have studied clinical judgement from a problem–solving perspective. The differences between decision-making and problem-solving approaches are largely historical and methodological. While decision-making research is based on examining the way in which judgement deviates from a rational standard, the problem-

solving approach recognises the expert practitioner as a gold standard (Patel et al. 2002). Despite these differences, studies of clinical judgement making using problem-solving tend to mirror the two system approaches (Benner 1982; Dreyfus and Dreyfus 1986). Studies of the diagnostic decision-making of doctors have suggested that clinicians who are inexperienced, or experienced clinicians faced with a complex or unfamiliar task, tend to use an analytical mode of thinking known as hypothetico-deductive reasoning, or information processing (Elstein and Schwarz 2002). Although described using different terminology, this is an exemplar of System Two thinking, involving the collection of information or cues, which are then used to generate one or more hypotheses. This information is then weighed against the hypotheses for correctness of fit until a particular hypothesis can explain the information collected. It has been suggested that this process transforms the original diagnostic problem (what is wrong with this patient?) into a series of better-defined problems (Elstein and Schwarz 2002). In contrast, experienced clinicians, faced with familiar diagnostic tasks, tend to use a form of intuitive judgement known as pattern recognition, where they automatically retrieve the diagnosis from a network of stored knowledge (Elstein et al. 1990).

The novice to expert approach (Dreyfus and Dreyfus 1986; Benner 1982) has gained widespread acceptance in nursing and midwifery, with its emphasis on intuition, craft knowledge and in particular the unknowable ‘art’ of the expert practitioner (English 1993; Davis-Floyd and Davis 1997; Gaskin 2002). The theory suggests that the practitioner must go through a series of five ascending

stages of proficiency in becoming an expert. These are novice, advanced beginner, competent, proficient and expert. This contrasts with the dual systems theories and the information processing theory, described above, which suggest that it is the task or task characteristics which determine the mode of cognition used. The novice to expert approach suggests that the novice must think analytically and apply explicit rules. Progressing through the stages of proficiency, the use of intuition increases exponentially, until as an expert, intuition is the principal mode of cognition used. Dreyfus and Dreyfus (1986) argue against the use of decision rules suggesting that they reduce the opportunity for the novice to develop expertise. It is not clear, however, at what point the thinking style of the emerging expert changes from analysis to intuition. In addition, as Benner and Tanner (1987) define intuition as 'understanding without rationale' it is not evident how an expert practitioner, practicing predominantly intuitively would be able to pass on his/her expertise, to a more inexperienced colleague or defend his /her clinical judgements if required to do so. Thus, because intuition is the domain of the expert and cannot be understood, except by the expert, it cannot be criticised for fear of the critic being labelled incompetent, in an emperor's new clothes type of dilemma.

3.7. Decision support

Despite the widely accepted notion that experienced clinicians will make optimum judgements predominantly using intuition, there is considerable evidence that this is not the case. Studies comparing the use of clinical judgement alone with decision support methods have consistently found that

judgements using decision support almost always performed better. While clinical judgement is based on the informal assembling of data by the clinician (Grove et al. 2000), decision support tools are based on the principle that there is an empirically established link between the data used and the event to be predicted (Dawes et al. 2002). They include statistical prediction rules and actuarial methods, and have a wide range of possible formats including algorithms (a step-by-step problem-solving process, expressed as a flow chart) decision trees (a diagram depicting decision options and possible consequences) and computerised decision support systems. Kawamoto et al. (2005) define decision support as

> any electronic or non-electronic system designed to aid directly in clinical decision-making, in which characteristics of individual patients are used to generate patient specific assessments or recommendations that are then presented to clinicians for consideration. (p765)

Thus to be considered decision support a tool must use data collected from an individual to provide guidance on that particular case, rather than give general guidance for management of a particular condition. Using this definition, guidelines, protocols and pathways would not be considered to be decision support tools, although they could contain decision support. It may be helpful at this point to make a distinction between decision support, as described above, and decision aids, which in healthcare literature predominantly refer to patient

decision-making or shared professional / patient decision-making (O'Connor and Jacobsen 2003). A considerable body of literature exists on decision aids (Bekker et al. 1999) which has not been included in this overview, although it has been referred to where necessary.

3.7.1. Studies evaluating decision support

The results of studies of decision support systems compared to unaided clinical judgement are summarised in six key papers (Meehl 1954; Grove et al. 2000; Dawes et al. 2002; Garg et al. 2005; Kawamoto et al. 2005; Randell et al. 2007). These are discussed in turn.

Meehl (1954), in the earliest collection of results of studies of expert versus actuarial judgements, identified 20 reports in which expert prediction was compared with some form of statistical prediction. The study topics were not exclusively clinical in nature including predictions about success in training or education (e.g. college students and naval cadets), behaviour of offenders and recovery from major psychosis. These studies originated from 1930 –1940s, varied in quality and in the amount of data presented and in the form of actuarial judgement compared. Nevertheless, in almost half of the studies included, actuarial judgements were found to be superior to the judgement of experts, in half, actuarial judgements performed equally well and in only one study did expert judgement out perform statistical prediction.

Grove et al. (2000) conducted a meta-analysis of studies comparing clinical judgements with mechanical prediction. Mechanical prediction included statistical prediction rules, actuarial prediction and computer-based algorithms. The analysis included 136 individual studies and found that in 47% mechanical prediction was superior, in a further 47%, mechanical prediction and clinical judgements produced the same results and in only eight studies (6%) clinical judgement was superior. The findings were consistent across a wide range of topic areas, including studies where clinicians had a range of level of experience. A limitation of this meta-analysis was the inconsistent quality of some of the included reports. However, the consistency of the results across such a large number of studies supports the conclusion that mechanical prediction is at least as successful as, and frequently more successful than, clinical judgement alone.

Dawes et al. (2002) summarised the results of a number of studies in which clinicians' judgement was compared with mathematical decision rules. For example, Goldberg (1968) developed a mathematical rule for interpretation of the Minnesota Multiphasic Personality Inventory (MMPI). Clinicians' performance using this rule was then compared with the performance of clinicians' who were unassisted in interpreting the MMPI. The study found that the mathematical rule consistently outperformed clinicians' unassisted judgements. In a second study Goldberg (1970) developed linear rules based on the judgements of clinicians. This study found that these rule based judgements outperformed the judgements of the clinicians on which they were

based, suggesting that consistency is a factor in the success of such decision rules. Other studies explored decision rules in relation to a range of clinician experience (Leli and Filskov 1984), clinician training in the use of decision rules and clinicians given the option of using the decision rule (Dawes et al. 2002). These studies consistently found that decision rule aided judgements outperformed clinicians' judgement alone.

Kawamoto et al. (2005) conducted a systematic review of RCTs of decision support systems with the aim of identifying the specific features of successful systems. The criteria for inclusion in the review were that the systems had to address practice in a real clinical setting and be used by clinicians in providing direct patient care. Seventy RCTs were included, most were computer-based (34%) with non-electronic systems comprising 26%. The review identified four key features of successful decision support systems. These were;

- Providing decision support as part of the clinicians' workflow
- Providing decision support at the time and place of decision-making
- Providing a recommendation rather than just an assessment
- Using a computer-based system.

Overall, decision support systems were found to improve clinical practice in 68% of included studies. However, this increased to 94% for systems which included all four features identified as most crucial to success. The review

found that where clinicians were required to seek information the system was less likely to be effective. Interestingly, the authors comment that they were unable to report on the effectiveness of decision support systems that aimed to directly improve patient outcomes (rather than improving the process of care) because very few studies reported these outcomes.

Garg et al. (2005) reviewed controlled trials of computerised clinical decision support systems. One hundred trials were reported including studies of diagnostic systems, reminder systems, disease management and drug prescribing systems. The review focussed on improvement in practitioner performance, improved patient outcome and factors contributing to successful systems. The review found that 64% of studies reported improvement in practitioner performance using the computerised decision support systems. However, the results relating to improvement in patient outcome were more equivocal. As with the findings of Kawamoto et al. (2005), a reduced number of studies (52%) reported the effects of decision support on clinical outcomes for patients. Further, most were underpowered to report these outcomes. Overall, only seven studies reported improved clinical outcomes as a result of computerised clinical decision support. The review found that systems in which practitioners were automatically prompted to use the system were more successful than systems where the practitioner had to actively seek decision support, and this agrees with the findings of Kawamoto et al. (2005). While reported barriers to success included; practitioners failing to use the system, poor integration into clinician workflow and practitioners refusal to accept

computer decision support. The review concludes that further research is required which is statistically powered to determine the effects of such systems on clinical outcomes. In addition, it concludes that there are complex factors associated with the successful implementation of computerised decision support systems and that in some situations cheaper and more effective non-computerised systems may be equally or more effective in improving clinical care.

Randell et al. (2007) reviewed experimental and quasi-experimental studies which evaluated the effects of computerised decision support systems in nursing practice. Eight studies of mixed methodological quality were included in this review. Three studies compared the performance of nurses using computerised decision support with nurses using clinical protocols or unaided clinical judgement. The remaining five studies compared nurses using computerised decision support with other professionals (for example doctors) not using such support. The results were equivocal. Three studies found that patient outcomes were improved where computerised decision support was used, four studies found no difference in patient outcomes and one study found that patient outcomes were poorer where computerised decision support was used. The heterogeneous nature of these studies makes drawing conclusions from the results difficult. Only three of the studies compared nurses' performance with and without decision support while the remainder compared nurses' performance using decision support with other professionals including doctors whose judgement performance may have been expected to be different

to that of the nurses regardless of decision support. The authors conclude that studies of more consistent methodological quality are required to evaluate the efficacy of computerised decision support in nursing.

A limitation of all of these reviews is that they only report the results of published studies; this means that there is likely to be publication bias against studies which found negative or neutral results. Nevertheless, there is a large body of research evidence to support the notion that decision aided judgements perform at least as well or better, than clinicians' un-aided judgements.

3.8. Why is decision support effective?

Several authors suggest reasons why decision support tools outperform clinical judgement. Meehl (1954) and Grove et al. (2000) suggest that clinicians rarely receive feedback on the outcomes of their judgements and decisions and therefore may be overconfident in their judgement accuracy. They also propose that some clinical judgements represent a self-fulfilling prophesy. For example, where a woman is admitted to a labour ward with a diagnosis of active labour it is more likely that, if she does not subsequently progress as expected, a diagnosis of slow progress will be made and her labour will be augmented with oxytoxics, rather than the original diagnostic judgement revised.

Dawes et al. (2002) suggest that it is the consistency of decision support tools which makes them effective. Given the same set of data the same judgement will be produced on each occasion, while even experienced clinicians' judgements will be characterised by random fluctuation which reduces reliability. This suggestion is supported by the findings of a study of the diagnostic judgements of nurse practitioners (Rosenthal et al. 1992) which showed that a linear model with as few as three cues performed as well as the nurse practitioners, suggesting that it was the inconsistent use of informational cues rather than lack of knowledge of the appropriate cues which led to diagnostic inaccuracy. In addition, the heuristics and biases approach described above (Kahneman et al 1982), suggests that people are not good at assessing probabilities, they ignore the base rate in making estimates of the likelihood of conditions, frequently seek mainly information which confirms their initial hypothesis and have overconfidence in clinical judgements based on personal experience.

Although decision support has been found to improve clinical judgement, a number of studies have found that it is underused by clinicians (Garg et al. 2005; Kawamoto et al. 2005). Recent studies of nurses' decision-making (Thompson et al. 2004; McCaughan et al. 2005) have found that nurses, faced with uncertainty in clinical decision-making, rely on personal experience or advice from colleagues, rather than text or electronic information or support. Thompson et al. (2004) in an observational study of nurse decision-making in an acute care setting, found that in 180 hours of observation, only two forms of

text based information were used, these were local protocols or guidelines (used four times) and the British National Formulary (used 50 times). It may be that decision support tools are not available for the sort of judgements and decisions routinely made by nurses and midwives, however, even where such systems are available they are often underused. For example, in a study of a decision support tool for prediction of acute ischaemic heart disease which had been found to reduce the false positive diagnosis rate from 71% to 0% the subsequent utilisation of the tool by clinicians was only 2.8% (Corey and Merenstein 1987). A number of possible reasons for this have been suggested. In particular, that decision support mediates against individuality of care (Tavakoli et al. 2000; Trinder 2000), and that it undermines the clinical skills of the practitioner (McCaughan et al. 2002; Tavakoli et al. 2000). It has also been suggested that use of decision support tools may undermine the clinical credibility of practitioners. A recent vignette-based study (Arkes et al. 2007) found that the diagnostic ability of doctors who used decision support was rated as lower than those who used clinical judgement alone. Interestingly in this study the ability of a doctor who used a decision support tool and then ignored its recommendation, was judged to have even lower diagnostic ability than those who used an aid and adhered to its recommendation. It appears that, despite the significant evidence for the effectiveness of decision support tools, clinicians and even the public may be sceptical of their value.

3.9. Conclusion

The premise of this thesis is that diagnosis of labour is a judgement made under conditions of uncertainty in which there is limited time in which to make

the judgement, emotional pressure, a large number of cues (some of limited salience (Burvill 2002)), and some uncertainty of outcome. This may predispose the midwife to use an intuitive style of judgement which is prone to error (Kahneman et al. 1982). Altering the judgement task, by introducing a decision support tool would provide structure to the judgement task, reduce the number of cues to be considered and increase the salience of cues. This has the potential to induce a more analytic judgement style, prompting consistency of collecting and processing of relevant information (Hamm 1988), thus reducing judgement error.

There is considerable evidence that the use of decision support tools may improve clinicians' judgements (Grove et al. 2000; Dawes et al. 2002; Kawamoto et al. 2005; Garg et al. 2005), although there are fewer studies which have reported on improved clinical outcomes. Factors which have been associated with successful systems are: providing decision support as part of the clinicians' workflow, providing decision support at the time and place of decision-making, providing a recommendation rather than just an assessment and using a computer-based system (Kawamoto et al. 2005). However, it has been suggested that in some situations cheaper, non-computerised systems may be equally or more effective in improving clinical care (Garg et al. 2005).

Although many maternity units in the UK have computerised patient information systems, they are not yet used universally and may not be available at the point of decision-making (this would be at the place where labour assessment is

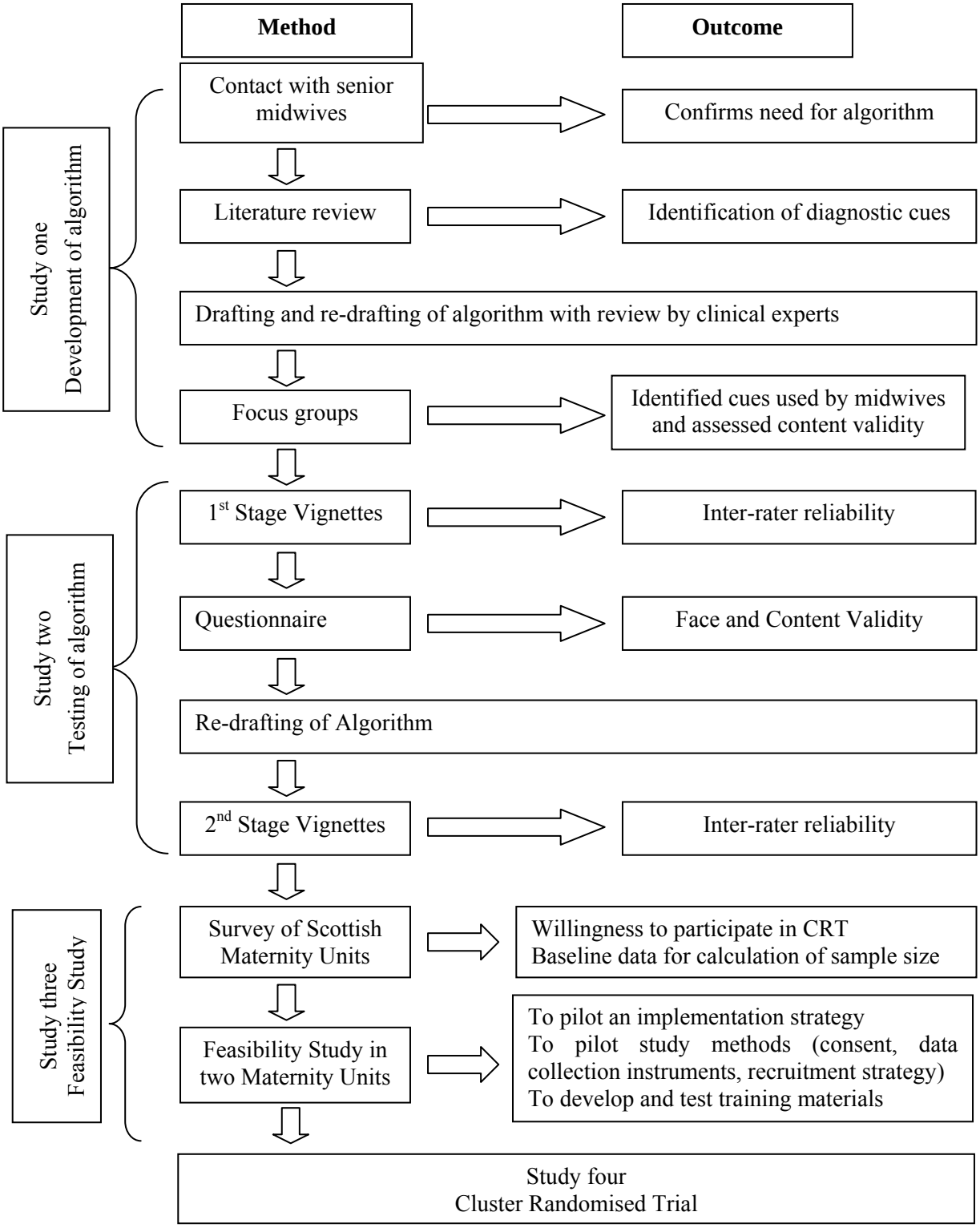
undertaken in the case of diagnosis of labour). For this study it was therefore, decided to develop a paper based decision support tool, in the form of an algorithm, for diagnosis of active labour. This algorithm would be an individual assessment tool (rather than a general guideline) available at the time of the admission assessment, as part of the midwives' workflow. The following chapter describes the development of the algorithm.

3.10. Overview of research methods

The research presented in this thesis comprises four studies which were conducted in a step-wise approach to developing and testing an algorithm for diagnosis of active labour in primiparous women.

At an early stage in the development of the study it was clear that a cluster randomised trial (CRT) would be the optimum method for testing the efficacy of the algorithm, however prior to the conduct of the trial a range of research methods (qualitative and quantitative) were used in developing and testing the algorithm and in conducting a feasibility study. This approach follows the format suggested by the MRC framework (2000) for development and evaluation of RCTs for complex healthcare interventions. An overview of the methods used for each study prior to the CRT is presented in Figure 2.

Figure 2 Overview of methods



CHAPTER 4: STUDY ONE: DEVELOPING AN ALGORITHM FOR DIAGNOSIS OF ACTIVE LABOUR IN PRIMIPAROUS WOMEN

4.1. Introduction

The review of the literature in chapter two identified that diagnosis of labour is often problematic, that many women are admitted to labour wards who are not yet in labour or who are in the latent phase (Ball and Washbrook 1996; Spiby et al. 2006a) and that these women are more likely to receive medical intervention than those admitted in active labour (Hemminki and Simukka 1986; Holmes et al. 2001). The challenge therefore appeared to be to more effectively discriminate between women who are in active labour and those who are not.

Chapter three provided an overview of decision-making theories, and in particular, the cognitive continuum theory (Hammond 1996) which suggests that when making decisions in situations of uncertainty people are likely to use an intuitive style of judgement which may be prone to error. This theory suggests that altering the judgement task by introducing a decision rule has the potential to induce a more analytic style of judgement, thereby reducing judgement error. Based on this theory, and on the evidence for the success of decision support tools (Meehl 1954; Grove et al. 2000; Dawes et al. 2002; Garg et al. 2005; Kawamoto et al. 2005; Randell et al. 2007), it was decided to develop a decision support tool, in the form of an algorithm, which aimed to support midwives to more effectively diagnose active labour in primiparous women. This chapter describes study one; the development of the algorithm.

4.2. Aim

The aim of study one was to develop an algorithm for diagnosis of active labour in primiparous women.

4.2.1. Objectives

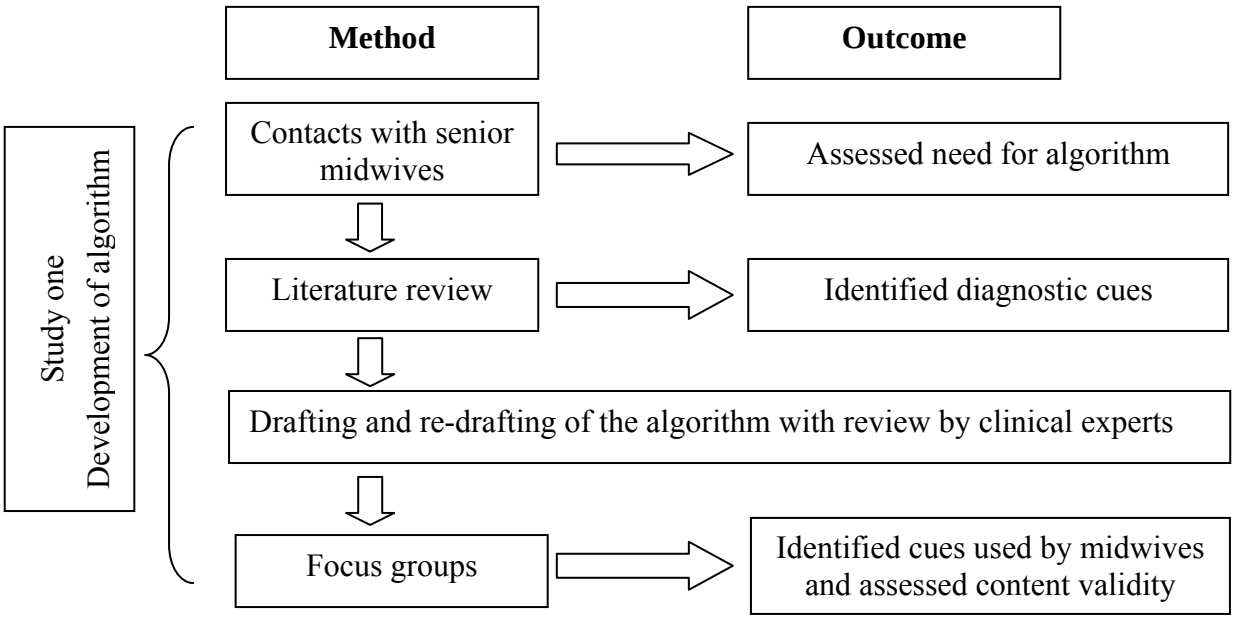
The objectives in study one were to:

- assess the need for an algorithm;
- identify informational cues for inclusion in the algorithm;
- develop the algorithm.

Methods

Figure 3 provides an overview of the methods used in study one. Documentation related to study one is presented in Appendix 2.

Figure 3 Study one methods



4.2.2. Assessing the need for an algorithm

O'Connor and Jacobsen (2003) recommend that a needs assessment should be carried out in the early stages of decision support development. Needs assessment, they propose, involves the compilation of evidence about the nature of the decision difficulty, numbers affected, availability of existing tools and demand for decision support. O'Connor and Jacobsen (2003) suggest that a variety of methods are applicable at this stage including theoretical development and clinical literature review as well as formal and informal surveys of stakeholders. Chapters two and three have described the review of decision-making theory and clinical literature which identified the degree of decision difficulty and the importance of the topic, including an estimate of numbers of women who are admitted to labour wards while not yet in active

labour or while in the latent phase. The remaining aspects of the needs assessment are addressed in this chapter.

Information about the availability of existing decision support tools and potential demand for decision support was gathered by means of informal telephone contacts with senior midwives and midwife managers in Scotland. These were carried out in the development stage of the study, prior to proposal development, with the purpose of gauging the need for the algorithm. The heads of midwifery or labour ward managers in fifteen consultant led maternity units were contacted. After a brief explanation of the purpose of the proposed study the conversation focused on the following questions: Had the unit in which they worked experienced problems associated with early admission of women to the labour ward? In their opinion would a decision support tool be useful? Was such a tool or a labour ward admission guideline currently in use? Their comments were recorded in note form.

All the senior midwives contacted said that their units had experienced the problem of women being admitted to the labour ward while not yet in active labour. All expressed support for the development of a decision support tool and interest in becoming involved with the proposed study. One of the managers reported that their unit was using a guideline for admission to an on-site birth unit. However, that guideline merely suggested that women admitted should be in active labour, but did not define it. The responses confirmed that none of the units were using an algorithm or guideline for the diagnosis of labour at that time.

4.2.3. Identification for informational cues for inclusion in the algorithm

Informational cues for inclusion in the algorithm were identified as follows: a literature review was conducted; the algorithm was then drafted and reviewed by a small group of clinical midwives and an obstetrician. Focus groups were then used to identify the cues which midwives reported using to diagnose labour and to determine their order and importance. This allowed comparison to be made between the cues included in the algorithm and those used by the midwives and provided an assessment of the content validity of the algorithm.

4.2.3.1. Literature review

The literature review was informed by a preliminary review of medical and midwifery texts. Key medical databases Medline, CINAHL and the Cochrane Library were then searched using the following search terms in the title or abstract: diagnosis of labour (labor), onset of labour, labour and active. The choice of database was informed by the fact that the focus of the research was a clinical issue, namely the diagnosis of labour, and it was felt that other databases (such as Embase) would be unlikely to extend the literature base. The Cochrane Library was determined to be a pertinent source of literature on RCTs, since it also includes the Dare database and HTA assessments. The search was limited to research papers published over the last 20 years; only papers which specified criteria for labour diagnosis were included.

Nine studies were identified (O'Driscoll et al. 1984; Turner et al. 1986; Boylan et al. 1991; Lopez-Zeno et al. 1992; Frigoletto et al. 1995; McNiven et al. 1998; Sadler et al. 2000; Burvill 2002; Janssen et al. 2003). These studies are

described in chapter two, the labour diagnosis criteria included in each are summarised in Table 2.

Table 2 Research studies identifying explicit criteria for diagnosis of labour

Included Studies	Design	Aim	Criteria for diagnosis of labour
O'Driscoll et al. 1984	Case series	A report on outcome of cases of dystocia	Painful uterine contractions in association with either bloody show, spontaneous rupture of membranes or complete effacement of cervix
Turner et al. 1986	Quasi-experimental	To evaluate an active management of labour package	Painful contractions accompanied by either; cervical dilatation with effacement, spontaneous rupture of membranes, show
Boylan et al. 1991	Before & After	To evaluate whether active management of labour would reduce the incidence of CS for dystocia	Regular painful contractions (at least one in 10 minutes) with at least 80% cervical effacement and 1 cm dilatation. Show or spontaneous rupture of membranes supports diagnosis.
Lopez-Zeno et al. 1992	RCT	To evaluate whether active management of labour would reduce incidence of CS	Regular painful contractions (at least one in five minutes) in association with complete cervical effacement or spontaneous rupture of membranes.
Frigoletto et al. 1995	RCT	To evaluate an active management of labour package	Painful contractions accompanied by effacement of at least 80%, show or spontaneous rupture of membranes
McNiven 1998	RCT	To evaluate explicit criteria for diagnosis of labour	Painful contractions Cervical dilatation greater than 3cm
Sadler et al. 2000	RCT	To evaluate an active management of labour package	Regular painful contractions (one in five minutes lasting 40 seconds) accompanied by either spontaneous rupture of membranes or full cervical effacement and dilatation of at least 2 cm
Burvill 2002	Qualitative	Exploration of midwifery diagnosis of labour	Describes: 1. Reactions of the woman (breathing, conversation, mood, energy and movement) 2. External signs (show, appearance of a red line between the buttocks, visual contractions, presenting part engaged) 3. Internal signs (cervical dilation of 3 or 4 cm, bulging membranes)
Janssen et al. 2003	RCT	Compared early labour assessment at home or by telephone triage	Regular contractions (at least 2 in 10 minutes) with cervical dilatation of at least 3cm

4.2.4. Drafting the algorithm

The algorithm was initially drafted with reference to the cues identified in the literature search, only physical cues, which could be objectively assessed were included. Development was based on the precept that it would be applied in a face-to-face consultation between a midwife and a woman and that prior knowledge of the woman by the midwife would not be a prerequisite for its use. It comprised three levels. Level one established the appropriateness of the tool for the specific population, i.e. a primiparous woman who had experienced an uncomplicated pregnancy (the term 'primigravid' was initially used and later altered to the correct term). Level two prompted a general physical assessment, including temperature, pulse and blood pressure. Level three presented, in a stepwise fashion, key informational cues required for the diagnosis of labour. As only one study had specifically evaluated the efficacy of specific cues, the inclusion of cues was based pragmatically on the frequency of their reporting rather than empirical evidence for their efficacy. Cues included were:

- Presence of abdominal pain (at least moderate), associated with regular uterine contractions (three in ten minutes), all of the studies included the presence of painful uterine contractions as the primary cue for diagnosis of labour.

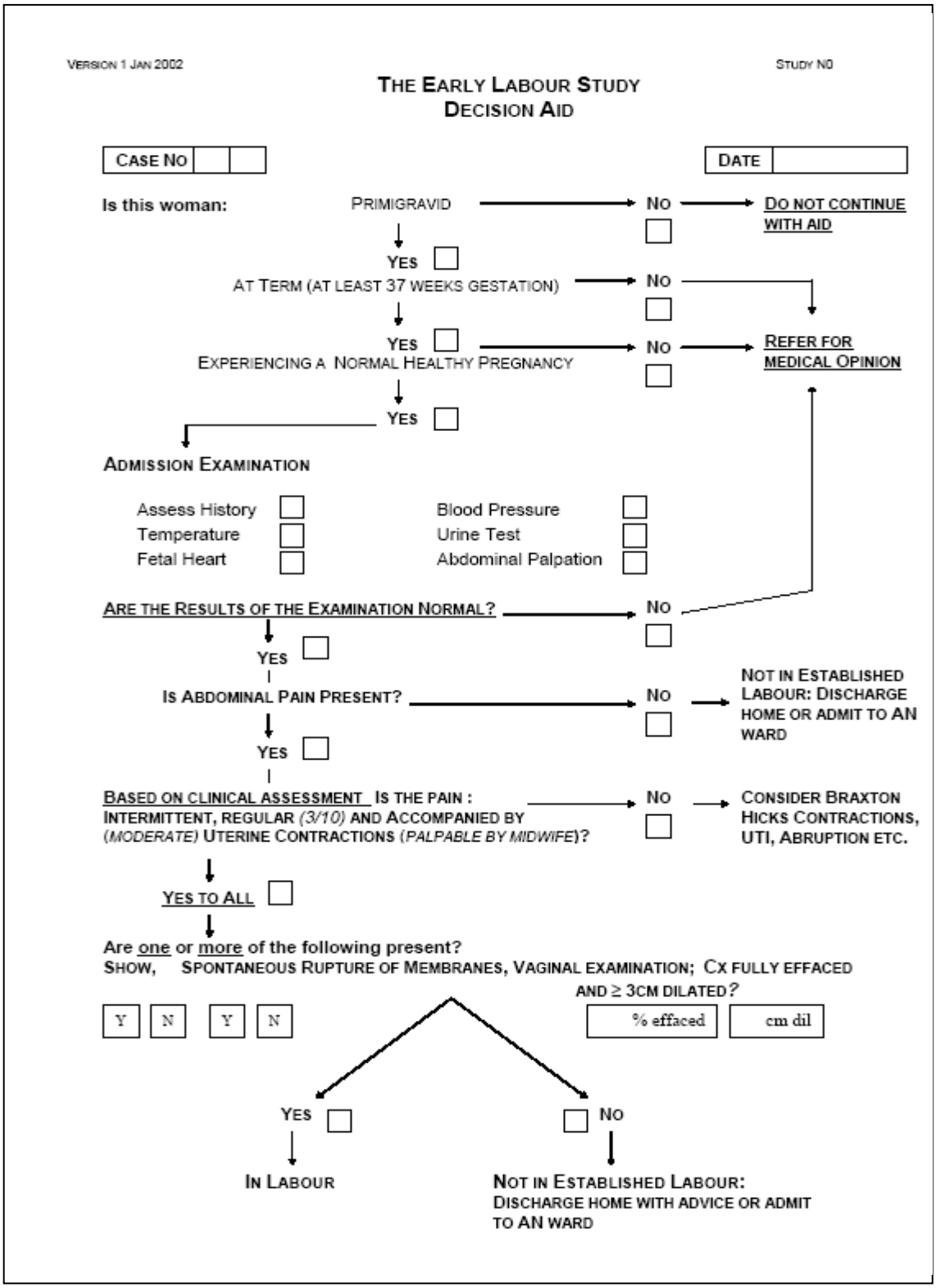
- Cervical dilatation (≥ 3cm and fully effaced): all studies included cervical dilatation, ranging from one to three centimetres, or effacement of the cervix, ranging from 'effacing' to complete effacement.

- Spontaneous rupture of membranes (SRM): five studies included spontaneous rupture of membranes (O'Driscoll et al. 1984; Turner et al. 1986; Lopez-Zeno et al. 1992; Frigoletto et al. 1995; Sadler et al. 2000) and one (Burvill 2002) included assessment of the condition of the membranes but did not specify SRM.

- Show: the inclusion of show was equivocal in the literature, however four studies included the presence of show (O'Driscoll et al. 1984; Turner et al. 1986; Frigoletto et al. 1995; Burvill 2002) and this was considered sufficient to justify its inclusion in the algorithm.

Using the algorithm, active labour would be diagnosed where, based on the clinical assessment of the midwife, regular painful contractions (as described above) were associated with at least one of the remaining cues. Thus, if regular painful contractions were associated with SRM, then vaginal examination would not be required solely for the diagnosis of active labour. The algorithm was reviewed by experienced clinical midwives and an obstetrician who were not otherwise associated with the study. The first draft of the algorithm is shown in Figure 4. This draft was entitled 'The Early Labour Study Decision Aid' and was used in this form, in study two (chapter five).

Figure 4 The Early Labour Study Decision Aid



4.2.5. Focus groups

The objectives of the focus groups were:

- to determine what cues midwives report using in diagnosing active labour;
- to determine the order of cues used;
- to explore the process of midwives' judgements and decisions about diagnosis of labour.

4.2.5.1. Methods

Focus group interviews have been extensively used in market, social and health research (Jackson 1998). Central to the success of the method is the notion that judgements and decisions are made in a social context and that data are therefore also elicited from within a social context, where participants' views are considered in relation to others (Robinson 1999). They are commonly used in the early stages of research to explore areas of uncertainty. Focus groups comprise a group of participants who have been selected, usually purposively, in order to address a specific research question and the group is facilitated by a moderator. The key feature which distinguishes the focus group from other interview methods is that they aim to draw on interaction between participants, actively encouraging discussion of anecdotes and experiences (Webb and Kevern. 2001). The focus group aims to explore a specific set of issues; however, they have the potential to generate a diverse range of themes

and ideas, which may be less easily accessible in one-to-one interviews (Webb and Kevern 2001).

The strengths and weaknesses of focus groups have been widely discussed. They are considered to be a cost effective way of generating a large amount of qualitative data quickly and are particularly useful in the early stages of research (Robinson 1999). The group setting may enable participants to express views which they might be more reticent to express in the context of a one to one interview (Lane et al. 2001). There is potential for contemporaneous quality control within the group, as participants may confirm or refute each others comments thus providing an element of face validity (Lane et al. 2001; Webb and Kevern 2001). In addition, the moderator of the group is able to re-cap and seek clarification of the issues raised.

Although it has been suggested that there is potential for the moderators' own views to bias the data collected (Lane et al. 2001), an issue which should be acknowledged in any form of data collection (qualitative and quantitative), it is possible that the focus group design may mediate against this. While a participant in a one to one interview may be reluctant to contradict the interviewer, as a member of a group, they may feel more empowered to disagree. While the group interaction is the key advantage of focus groups there are potential disadvantages. Group discussion may be dominated by one participant or a small group of participants (Lane et al. 2001), while less articulate participants may be reluctant to contradict the stronger members.

Thus a minority opinion may appear to represent the group norm. There is also potential for conflict and issues over confidentiality to arise within the group (Robinson 1999).

Focus groups were chosen at this stage of the study because, in this development stage, it was a priority to gather data quickly. Most UK births take place in the context of hospital labour wards where there is interaction between midwives. It therefore seemed appropriate to gather data on midwives diagnosis of labour in the social context of a focus group and to draw on the potential for dynamic group interaction rather than the one to one interaction in an individual interview. A potential disadvantage was the effect that dominant midwives could have on the group, and this highlighted the importance of effective group facilitation.

4.2.5.2. Participants

The focus group participants were midwives working in a large urban maternity unit in the North of England. This location was chosen to minimise the potential for bias in the subsequent cluster trial, which would be conducted in Scotland. The target population was midwives either currently working in the labour ward or with recent labour ward experience. A seminar was held in the maternity unit, at which information about the study was presented. Thereafter, written information, an invitation to participate and a pre-paid envelope were distributed to midwives to return contact details if they wished to participate. Those who volunteered were invited to attend one of two focus groups. It was anticipated

that between six and eight midwives would participate in each group, and a gift voucher was given to each. It has been suggested that segmented sampling may be useful in avoiding the potential for inhibition of contributions where a hierarchy exists in a group (Morgan 1995). In this study, however, difficulty in arranging focus groups around midwives' shifts prevented this and therefore groups comprised midwives of different clinical grades.

4.2.5.3. Data collection

The focus groups lasted for approximately one hour. Midwives were given an explanation of the nature and purpose of the study and asked for consent to participation, including tape-recording of the session. A short questionnaire, collecting demographic details, was completed by each midwife prior to the group discussion.

The groups were facilitated by the author of this thesis (HC) assisted by a research assistant. Both groups were conducted in the same way. In order to encourage freedom of discussion the midwives were not shown the algorithm. Two trigger questions were asked: 'How do you decide whether a woman is in active labour?' and 'What factors or cues do you take into consideration when a woman is admitted in labour?' Participants were given a few minutes to think about the questions and to make notes if they wished, and were then asked to discuss the questions as a group. As the participants identified cues, these were written on index cards by the research assistant. The cards were left on the table during the discussion as reminders for participants and were used in

re-capping by the facilitator. Participants were also encouraged to use the cards to clarify the order in which they identified cues in carrying out labour assessment. Subsequent comparison of the cues (and the order of cues), identified by the midwives and those contained in the algorithm provided evidence for the content validity of the algorithm.

4.2.5.4. Ethical considerations

Ethics approval was granted by the Local Research Ethics Committee (ref 2002/208) (Appendix 2) and permission for access by the Head of Midwifery. Before starting, the groups agreed that all discussion would remain confidential. It was stressed that quotations would be anonymous and the hospital would not be identified in reports or publications.

4.2.5.5. Data analysis

The groups were tape-recorded and transcribed verbatim by the author of this thesis (HC); indicators of group interaction, e.g. laughter, group agreement or dissent were also transcribed. There was a great deal of laughter, and dissent was often expressed in the form, 'I don't know about the rest of you but I always …..'. Frequent re-capping by the facilitator using the cue cards allowed views expressed by a minority to be explored and confirmed within the groups.

Data were analysed by hand (this was possible because of the relatively small size of the data set), using latent content analysis. In content analysis

categories and constructs are identified. Morse and Field (1996 p112) describe latent content analysis as a process by which

> Passages or paragraphs are reviewed in the context of the entire interview in order to identify and code the thrust or intent of the section and the significant meanings within the passage.

This, they suggest, allows both the overt intent and the underlying meaning of the participants to be included. Because of the potential for bias due to the subjective nature of latent content analysis, analysis was carried out independently by the author of this thesis and two supervisors in order to identify the main categories and themes which emerged from the discussion. Categories were compared, and discussed until consensus was reached on the themes emerging. There was remarkable similarity in the themes identified. Discussion primarily concerned the naming of themes (e.g. 'physical signs' versus 'clinical signs') and the placement of some cues which could appear in more than one theme. The placement of cues was most challenging within the category of institutional factors, and was resolved by agreeing on the use of a broader theme (organisational factors).

A possible model of decision-making was developed based on the apparent relationship between the categories and themes. This was discussed with an experienced researcher independent of the focus group study; various alternative explanations for pathways of decision-making were suggested and discussed until consensus was reached.

4.2.5.6. Findings

Initially 17 midwives volunteered to participate in the study; however, ultimately 13 were able to participate due to difficulty in arranging meetings to fit around clinical commitments. Two focus groups were conducted; six midwives took part in the first and seven in the second. Table 3 describes the characteristics of the participants.

Table 3 Characteristics of participants

Characteristic	Group one (n=6)	Group two (n=7)
Years midwifery experience:		
< 5	3	1
6-10	1	3
> 11	2	3
Clinical grade:*		
G grade	0	3
F grade	1	3
E/F grade	5	1
Current area of practice:		
Labour ward	3	6
Post-natal (time since labour ward experience)	2 (2 & 6 weeks)	0
Research (time since labour ward experience)	1 (4 months)	1 (2 years)
Qualification:		
Professional	2	2
First degree	4	4
Higher degree	0	1

* Clinical grades: The UK had a clinical grading system for midwives and nurses. Grade G – senior clinical midwife with team leader or ward management responsibility, Grade F – experienced clinical midwife, Grade E/F – junior staff midwife.

The same categories and themes were identified in both groups. The midwives described information cues which could be separated into two categories, those arising from the woman and those from the institution. The themes relating to the women were Physical signs, Distress and coping, Woman's expectations and Social factors, those, which related to the institution were Organisational factors, Midwifery care and Justifying actions (Figure 5).

Figure 5 Categories, themes and information cues

The themes are presented in turn, with quotes from the focus groups used to support the findings. Midwives in both groups initiated discussion about parity and agreed that diagnosis of labour was different in primiparous and multiparous women, where the woman's past experience was a dominant

feature. The subsequent discussion focussed primarily on the admission of primiparous women.

4.2.5.6.1. The woman

Midwives reported using a series of physical signs to diagnose labour. The first of these was the woman's appearance (her demeanour), which provided a physical sign of whether she was in pain and her response to the pain:

I look at her first, you think, what does she look like and what is her state? Is she actually uncomfortable, has she walked up crying or smiling?

They then described a series of cues which built up to confirm or refute their first impression. The presence of painful contractions was seen as essential to a diagnosis of labour. Where these were absent a differential diagnosis was sought:

You're looking for other symptoms, anyway. You'd be asking about urinary frequency or pain passing urine, because sometimes labour symptoms can be the same as urinary tract infection. So if somebody comes in in pain, but they're not having contractions, you'd be thinking, 'Do they have a urinary tract infection or a bleed?'

Strength, frequency and regularity of contractions were assessed:

Her contractions, how frequently they're coming. I'd palpate the contractions to see how strong they are and how long they're lasting.

Spontaneous rupture of membranes, which in the presence of contractions was indicative of labour, was an important management cue in the absence of contractions:

If they're not contracting (and have ruptured membranes) it impacts on your management, but it doesn't impact on what you're saying about labour.

The appearance of show was considered to be one of the signs, which, although important to the woman, was not considered to be so by any of the midwives:

In labour, when you look at a lady's pad and you see show, you think, 'Oh things are changing with the cervix', but not when she's admitted.

In contrast, the vaginal examination (VE) was considered to be very important in establishing whether the woman was in labour, in particular, before sending her home. However, as with contractions, an aggregation of individual findings often took place. For example, an assessment of the position of the cervix, along with its constituency, application, degree of effacement and dilatation, was required. The midwives agreed that considering a combination of factors was essential, in particular groups of cues in the presence of contractions and all aspects of the vaginal examination rather than dilatation alone:

If someone was contracting regularly, the cervix was dilating, she'd ruptured her membranes, you'd think, 'Great - things are going nicely.'

although some of the midwives did appear to place particular weight on one or two significant cues:

If it's (the cervix) posterior and not effaced, she's not in labour.

Overall, the midwife's assessment of the woman's history was important. In particular, women who had repeated admissions in false labour were a cause of concern to a number of the midwives:

Sometimes they've been in and out thinking they're in labour, and you're thinking, 'Is there something wrong? Or, should we be sending her home?' Because obviously she's not dealing well with what's going on.

A number of participants highlighted that the diagnosis of labour could best be seen in retrospect, or in relation to the passage of time:

It's when her contractions are becoming more regular, they're becoming more painful, they're becoming stronger - that's when she's in labour.

The physical cues used by midwives are summarised in Table 4.

Table 4 Cues described by midwives for diagnosis of labour

Cues	
General	**Specific**
General appearance	Physical signs of pain, response to pain
History	Parity and gestation
Uterine contractions	Strength, regularity and frequency
Spontaneous rupture of membranes	*In the presence of contractions* considered to be indicative of labour. In the absence of contractions, important in relation to the woman's management
Show	Considered to be important to women but not to midwives, may indicate progress in labour in the presence of other signs
Vaginal examination	The following elements were considered equally important in the presence of contractions; cervical effacement, cervical application (well applied), cervical dilatation (at least 3cms)

Midwives reported that, at the same time as assessing physical cues, they had to consider a number of other factors which led to a decision on the most appropriate management. The woman's level of distress and how well she was coping with the pain she was experiencing were important, this included fear and need for reassurance even if she was not in labour:

Sometimes you'll have a lady who comes in, cervix only 50% effaced, maybe one centimetre, quite posterior, but she's so distressed you just couldn't possibly send her home. So you would keep her in, not because she's in labour but she's not coping, she needs reassurance.

One theme which appeared to occur both as a physical cue and within the theme of distress and coping was the woman's appearance. This seemed to provide information about her pain and ability to cope with what she was experiencing. Distress and coping were strongly related to what the woman's expectations of labour were, and how well prepared for labour she was:

You realise that they have not a clue of what to expect and therefore they're scared.

Participants felt that many of the women had unrealistic expectations and that depiction of labour in the media, in particular in soap operas, was misleading. They suggested that antenatal education did not fully prepare women for what to expect in early labour, e.g. pain arising from cervical effacement. Women attend hospital feeling that they were in labour, and a number of midwives had experienced conflict between their own clinical findings and the expectations of the woman, or her family:

I've often given women sedation against my better judgement because her mother has insisted that that's what she wants, and I haven't felt that I was managing that the most effective way.

A very important aspect was that of the social factors involved, e.g. the support of the woman's family, her partner's level of anxiety or that of her mother:

I always take a look at the partner, 'cos occasionally you're thinking ahead - if this girl goes home is she likely to cope with the support she's got…or come straight back in?

The distance or means of transport to the hospital was also important:

Sometimes if they've come in by ambulance and you're sending them home, then they're paying for a taxi to go home and then they're going to call an ambulance to come back in again.

4.2.5.6.2. The institution

As well as assessing the physical cues and other factors arising from the woman herself, the midwives had to work within the framework of the institution in which care was delivered. They had to negotiate a number of organisational factors, in particular pressures of workload including lack of beds and shortage of staff:

She may want to stay for the reassurance, and you are desperately trying to shove her out the door because you are just heaving at the seams and you've got nowhere to put her or no midwife to look after her.

Another organisational factor was the constraints imposed by clinical guidelines. Midwives felt that within these constraints they had limited options for the care of a woman who was not yet in active labour, while at the same time they had to protect her from a cascade of interventions:

*But you've got to work out, because we've got these guidelines, that she might be coping very well and you know she's in labour and she's managing really nicely. Oh, but she's been here four hours and she's still two centimetres, so you don't want her to run into the syntocinon and ARM (*artificial rupture of membranes*).*

An important aspect within this framework was the model of midwifery care. Assessment of the woman was more difficult as midwives did not know her beforehand, and lack of continuity meant that they had to make an 'on the spot' judgement of how she was coping; these judgements were often based on stereotypes. Midwives felt that models of care, which allowed midwives to get to know the woman prior to admission in labour reduced this problem:

When we talk to them on the phone we've never met them before, so you make your on- the-spot judgement. Whereas the team midwives, they know the ladies … so their advice is tailored, isn't it. While often we have to say, 'Come in'.

The need to justify their actions to others was a factor influencing midwives to send a woman home, and of particular concern was the midwife in charge during a shift. This appeared to be a fear of others' opinions and of being judged or blamed, and in some cases led midwives to provide care surreptitiously:

You often feel you have to justify your decision, (to?) to the midwife in charge.

I did hide a woman in the first stage room for four hours. I knew she was in labour but she was only one centimetre (dilated). Well, it was (names midwife), who would send people home unless they were pushing.

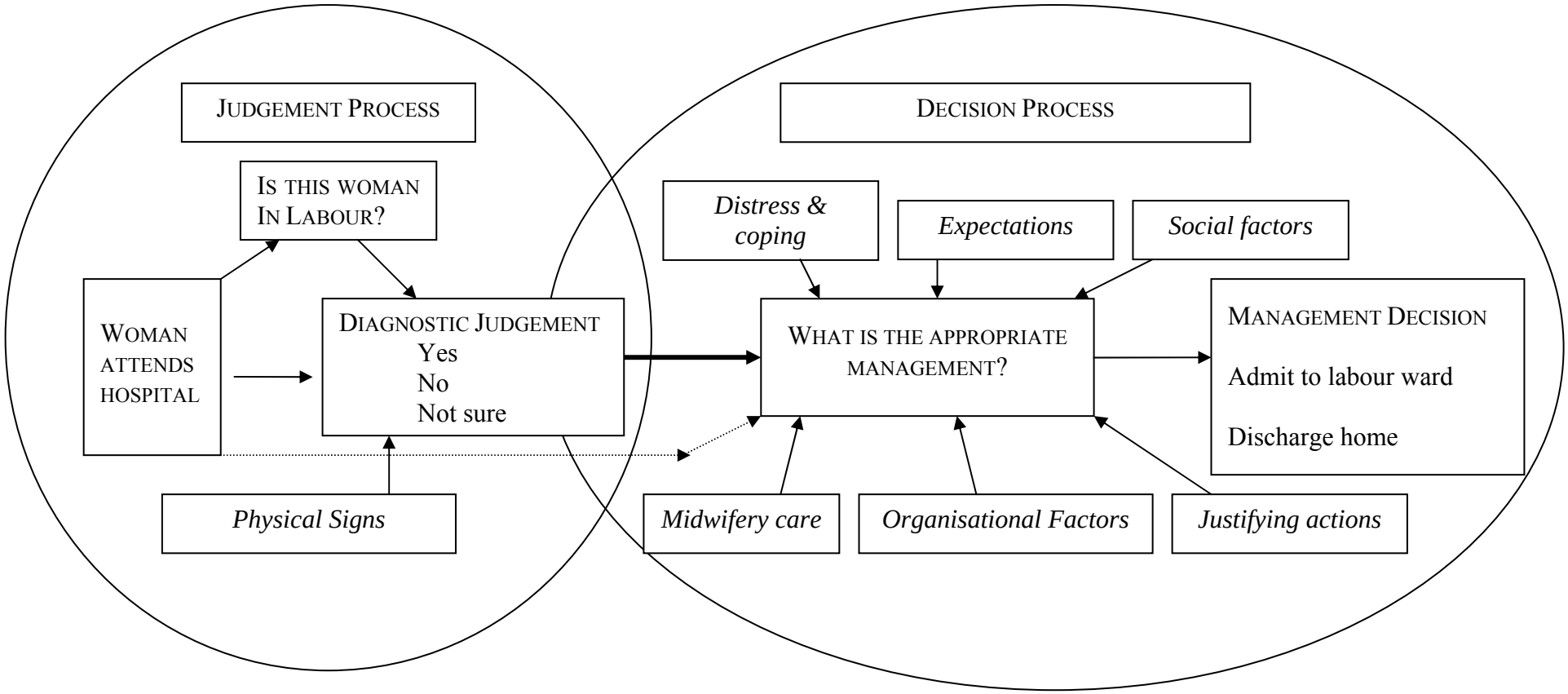
It appeared that the decision-making process could be divided into two distinct stages: the diagnostic judgement and the management decision:

To assess women, to make a professional decision, I think you can make that assessment quite early on. But your management, I think that could take another hour and a half.

4.2.5.6.3. Model of judgement and decision-making

The findings allow a model of judgement and decision-making to be suggested (Figure 6) which may assist in understanding the process of labour diagnosis and admission management by midwives. The midwife requires to make a diagnostic judgement (to answer the question, is this woman in labour?) and a management decision (to answer the question, what is the appropriate management for this woman?). The diagnostic judgement is usually made first, based on the physical signs (cues) of labour. Where a woman is in active labour, the management decision would be clear-cut. However, where the midwife's judgement is that the woman is not in labour (or not yet in active labour), the management decision would be made by considering a series of competing cues (including, but not exclusively based on, the diagnostic judgement), and in particular how the woman was coping, her expectations and those of her family as well as the requirements of the institution. Although the model suggests a predominantly sequential process, it acknowledges that the midwife may, in some cases, make a management decision before, or even in spite of, the diagnostic judgement, for example, where a woman is too distressed to be discharged, despite not being in labour.

Figure 6 Model of decision-making for diagnosis of labour



4.3. Discussion

4.3.1. Cues and order of cues used in diagnosing labour

The literature review included only studies which specified cues for diagnosis of labour. Most of the studies identified the same cues, referring to the O'Driscoll work on active management of labour (O'Driscoll et al. 1973), with some variation in particular in relation to cervical effacement (effacing, 80% or fully effaced). As only one study had evaluated efficacy of cues in improving clinical outcomes inclusion and exclusion of cues was based on frequency of inclusion rather than evidence of efficacy.

The focus groups indicated that midwives did describe using mainly physical cues in making a diagnostic judgement. The cues which they reported using (summarised in Table 4), were the same as those identified in the literature review and contained in the algorithm, this suggests that it had good content validity.

The midwives were able to rank the importance of cues and the order of their use. In particular, they placed high importance on the presence of regular, painful, uterine contractions which is the principal cue in the algorithm. Other signs were considered diagnostic of labour only in the presence of contractions and this suggests that the ordering of the cues in the algorithm was appropriate. However, the midwives also highly rated their first impression of the woman's appearance and demeanour. This is similar to the findings of Burvill (2002)

who included various aspect of the woman's appearance in a model of labour progress. However, cues such as these are elusive, they are difficult to include in an algorithm and have less salience with the process of labour than the physical cues. Further, while the algorithm gave equal rating to cervical dilatation, spontaneous rupture of membranes and show, the midwives felt that show had more importance to women than to their clinical assessment and a few midwives reported using cervical dilatation alone in diagnosing labour.

4.3.2. Process of judgement and decision-making

Within the judgement and decision-making literature a distinction is often made between a judgement, defined as an assessment of alternatives, and a decision, which is a choice between alternatives (Tversky and Kahneman 1974). The findings from the focus groups suggest that it is possible to make this distinction between the midwives' diagnostic judgements and their decisions about management. This is useful in that it permits different aspects of the judgement and decision-making process to be considered separately, and allows a hypothesis to be framed about the way in which these two aspects may interact.

First, considering the diagnostic judgement, the midwives placed considerable importance on their first impression of a woman's appearance; this is suggestive of intuitive thinking in which a number of cues are synthesised simultaneously. The woman's appearance gave information about the physical cue of pain as well as about her ability to cope with the pain she was

experiencing. However, as midwives did not have the opportunity to know the woman beforehand they felt restricted by the need to make an 'on-the-spot' judgement about a woman not known to them. Studies have identified culturally- bound behaviour in response to childbirth pain (Callister et al. 2003), and this might be expected to affect a woman's appearance; however, cultural factors were not raised in this study.

The findings also suggest that midwives aggregated cues, for example, within the vaginal examination (effacement, dilatation, application and position). This suggests use of an analytical mode of judgement. While clinical reasoning may appear to be more rational when discussed retrospectively (as in a focus group) than would be the case in the complex clinical situation, these findings suggest that midwives use both intuitive and analytical thinking in making the diagnosis of labour. Although, this supports the notion of the cognitive continuum theory (Hammond 1996; Hamm 1988) and dual process theory (Sloman 1996; Kahneman and Frederick 2002), these data do not provide sufficient evidence to determine whether midwives' intuitive thinking is corrected by analytical thinking or whether a 'middle ground' or quasi-rationality is being used (Hammond 1996).

An interesting finding was the interaction between the woman, midwife and institution which appeared strongly to influence the management decision. Although diagnosis of labour was made on the basis of physical cues, the management decision was not necessarily based only on that diagnosis. For

example, where a woman was definitely in labour then admission to labour ward was the normal decision. However, where a midwife judged that a woman was not in labour, or not yet in active labour, then negotiation was needed between clinical judgement, pressure from the woman seeking admission and pressure from the institution to keep her at home. This could lead to conflict which, the midwives reported, could result in sub-optimal management, e.g. giving sedation 'against her better judgement', 'hiding a woman' or midwives being unable to provide care for a woman who required it solely because she was not yet in active labour.

Other authors have explored the relationship between the diagnostic judgement and the treatment decision. The classic diagnosis/management model suggests that the clinician makes a diagnosis and, based on that diagnosis, a treatment decision (Barrows and Pickell 1991). However, Sorum et al. (2002) suggest that this may not always apply. They hypothesise that this sequence of judgement and choice is often violated in practice, where clinicians may decide on treatment before arriving at a diagnosis; they propose an alternative model of clinical decision-making in which the diagnostic judgement and treatment choice are made by means of independent, although largely simultaneous, processing of diagnostic and treatment cues.

The findings of the focus groups suggest that there is a predominantly sequential element in the relationship between the diagnostic judgement and the management decision. Midwives appeared to diagnose labour based on

physical cues before assessing aspects such as distress and coping and social factors (Figure 5). Thus, the diagnosis of labour became a factor (but not the only one) in the subsequent management decision. However, the findings suggest that this process is not clear-cut. The first cue assessed by a number of midwives was appearance, and this was common to both the physical cues (used in diagnosis) and distress and coping (used in management). This suggests that a simultaneous processing model, similar to that suggested by Sorum et al. (2002), may be used by some midwives or in certain circumstances.

Separating the elements of the judgement and decision-making process is useful in understanding why a superficially straightforward process may be problematic in practice. In addition, it identifies key points at which the introduction of decision support may improve the process and outcomes of judgements and decisions. The diagnostic judgement is one such key point where the introduction of the algorithm, has the potential to improve midwives' diagnostic judgement.

4.3.3. Limitations of focus group method

The use of focus groups in this study generated dynamic discussion and a wealth of data about midwives' experience of diagnosing labour, and this was a key objective of this development stage. However, the results are not generalisable as the sample was small and drawn from one UK maternity unit only. Therefore the findings may not be applicable to other geographical or

cultural contexts. The groups were similar in relation to themes and cues identified; this was unsurprising, as all midwives were working in the same maternity unit. In addition, these findings represent only the perspective of the midwives; although they speculated about the feelings of women, these data may not be considered to represent women's views. Most participants worked in rotation through different clinical areas; although four were not currently working in delivery areas, all but one had recent labour ward experience.

4.4. Summary

This chapter has described the process of developing an algorithm for diagnosis of active labour in primiparous women following an informal process of consultation with senior midwives throughout Scotland. The cues for inclusion in the algorithm were first derived from literature and the algorithm was then drafted and reviewed by a group of midwives and an obstetrician. Focus groups provided information about the cues used by midwives, the order of their use, and the process of judgement and decision-making.

The findings of the focus groups suggested that midwives use both intuitive and analytic thinking in diagnosing labour, and this fits with the cognitive continuum theory. The data from the groups also identified that midwives use the same cues as those identified in the literature and included in the algorithm, suggesting that it had good content validity. The algorithm was therefore not changed at this point. The following chapter describes the testing of the algorithm for reliability and further aspects of validity.

CHAPTER 5: STUDY TWO: TESTING THE ALGORITHM

5.1. Introduction

The first part of this thesis identified the association between the admission to labour wards of women who are not yet in active labour, or who are in the latent phase, and increased rates of medical intervention (chapter two). A review of theories of human judgement (chapter three) suggested that in making judgements in conditions of uncertainty people may rely on heuristics or intuitive judgements which are subject to increased error. In these situations the cognitive continuum theory (Hamm 1988; Hammond 1996) proposes that the introduction of a decision rule may reduce inconsistency in judgements and thus improve judgement quality.

Study one (chapter four) described the process of developing a decision support tool in the form of an algorithm. In this, and the following chapters, the process of testing the algorithm is reported.

5.2. Testing the algorithm

It is normal for any tool to be tested to determine whether it is fit for purpose. In healthcare, psychology and social sciences measurement and assessment tools would normally be tested for validity and reliability (Streiner and Norman 2003). While a number of aspects of validity have been defined the key element is determining whether the tool is useful. Reliability assesses the

degree to which the outcomes of the tool are reproducible within acceptable limits (Streiner and Norman 2003).

The algorithm required to be tested before it could be implemented in clinical practice. However, because it is a tool for assessing a physiological condition rather than a hypothetical construct, it required an approach to testing which was specifically tailored to its purpose. To be considered useful the algorithm had to meet specific criteria, namely, it had to be able to improve the diagnosis of active labour and it had to be understandable and acceptable to midwives. These criteria are discussed below.

5.3. Improving diagnosis of active labour

In order to improve diagnosis of active labour the algorithm would have to demonstrate three main properties. First, it would have to bring together the key diagnostic cues for active labour, at the correct level and in the right order. Second, as decision-making literature suggests that it is inconsistency of judgement that leads to error (Rosenthal et al. 1992; Dawes et al. 2002), the algorithm would have to promote consistency of midwives' judgements. Finally, the algorithm would have to produce evidence of improved clinical outcomes for women.

5.3.1. The diagnostic cues

The cues for diagnosis of active labour are physical and largely objective, although clinical judgement is required (for example, while cervical dilatation of at least three centimetres is an objective cue it cannot be measured with a ruler). The literature review (chapter two) identified that there was reasonable consensus regarding the cues which are required to diagnose active labour (O'Driscoll et al. 1984; Turner et al. 1986; Boylan et al. 1991; Lopez-Zeno et al 1992; Frigoletto et al. 1995; McNiven et al. 1998; Sadler et al. 2000; Burvill 2002; Janssen et al. 2003), although there was less agreement about the level of some cues (for example, how effaced the cervix would require to be). The algorithm was developed using these cues as described in chapter four. The content of the algorithm then required to be assessed, by seeking the subjective opinion of clinical experts, to ensure that the key diagnostic cues were included, and that these cues were ordered in the optimum way to facilitate the midwives' judgement.

5.3.2. Consistency of judgement

As well as containing the key diagnostic cues an important attribute of the algorithm would be that it would promote the consistency of midwives' judgements. In particular, that a number of midwives using the algorithm would come to the same judgement given the same information, within acceptable limits. This could be tested using paper based vignettes or scenarios of clinical cases in which the same information would be provided to each midwife.

5.3.3. Improvement of clinical outcomes

If the algorithm was found to meet the conditions described above, that is, if it contained the correct cues for diagnosis of active labour, and if these cues could be accurately and consistently recorded by midwives using the algorithm, then it would have the potential to be a useful diagnostic tool for active labour. However, for the algorithm to be useful in a clinically relevant sense, its impact on improving clinical outcomes for women would require to be tested. While the first two conditions could be tested using qualitative methods and paper based vignettes, this important aspect would require the algorithm to be tested in a clinical trial.

5.4. Acceptability to midwives

Finally, to be useful the algorithm would have to be used by midwives. A number of studies have highlighted clinicians' reluctance to use decision support (Thompson 2004; Garg et al. 2005; Kawamoto et al. 2005). The algorithm therefore required to be tested to ensure that it was in a form that midwives would recognise as having the potential to diagnose labour; for example, whether it looked right, whether the order of cues was consistent with the way in which midwives think about diagnosing labour, and whether it could be completed easily and with minimum effort. A final consideration was whether midwives would find the concept of an algorithm to be acceptable in principle.

The criteria discussed above are related to the traditional notions of validity and reliability in the following way. Assessing whether the correct cues, levels and order of cues have been included in the algorithm and whether midwives would recognise it as a tool which has the potential to diagnose labour maps onto the concept of face and content validity. Face validity relates to the whether a tool appears, superficially, to measure what it aims to, whether its purpose and relevance are self evident and essentially, whether it looks right. Content validity assesses the extent to which items within the tool adequately cover all aspects of the issue being addressed (Streiner and Norman 2003).

Assessment of consistency of judgement maps onto the notion of reliability which is used to describe the degree to which repeated measurements using a tool will produce the same result. Two types of reliability are commonly used; these are inter-rater reliability, the degree of agreement between different judges using the tool on a single occasion (Bowling 1991), and intra-rater reliability, a measure of the variation which occurs within one judge using the tool on different occasions (Streiner and Norman 2003).

While these elements of testing the algorithm (correct cues, acceptability and consistency) fit relatively well with the notion of face and content validity and reliability, the final element, that is whether the algorithm can improve clinical outcomes does not map so well onto traditional notions of validity. For example, two commonly used criteria are construct and criterion validity. Construct validity is used in situations where the measurement assesses a

hypothetical construct, for example, a theory which seeks to explain a behaviour or attitude (Bowling 1991). Although there is some debate about the meaning of labour (Gould 2000), it is generally considered to be a physiological state; therefore the notion of construct validity is not applicable. Criterion validity relates to the extent to which a tool predicts subsequent outcomes (Streiner and Norman 2003). The results obtained from the tool under scrutiny are compared with those obtained from an existing reference measure, a gold standard. Currently no such gold standard measure exists for diagnosis of active labour, therefore, a proxy measure of improvement in clinical management of labour has to be used (this is discussed in chapter six).

In this and subsequent chapters the testing of the algorithm is described, firstly using paper based modelling and questionnaires, thereafter in a feasibility study and finally in a clinical trial.

5.5. Aim

The aim of study two was to test the algorithm.

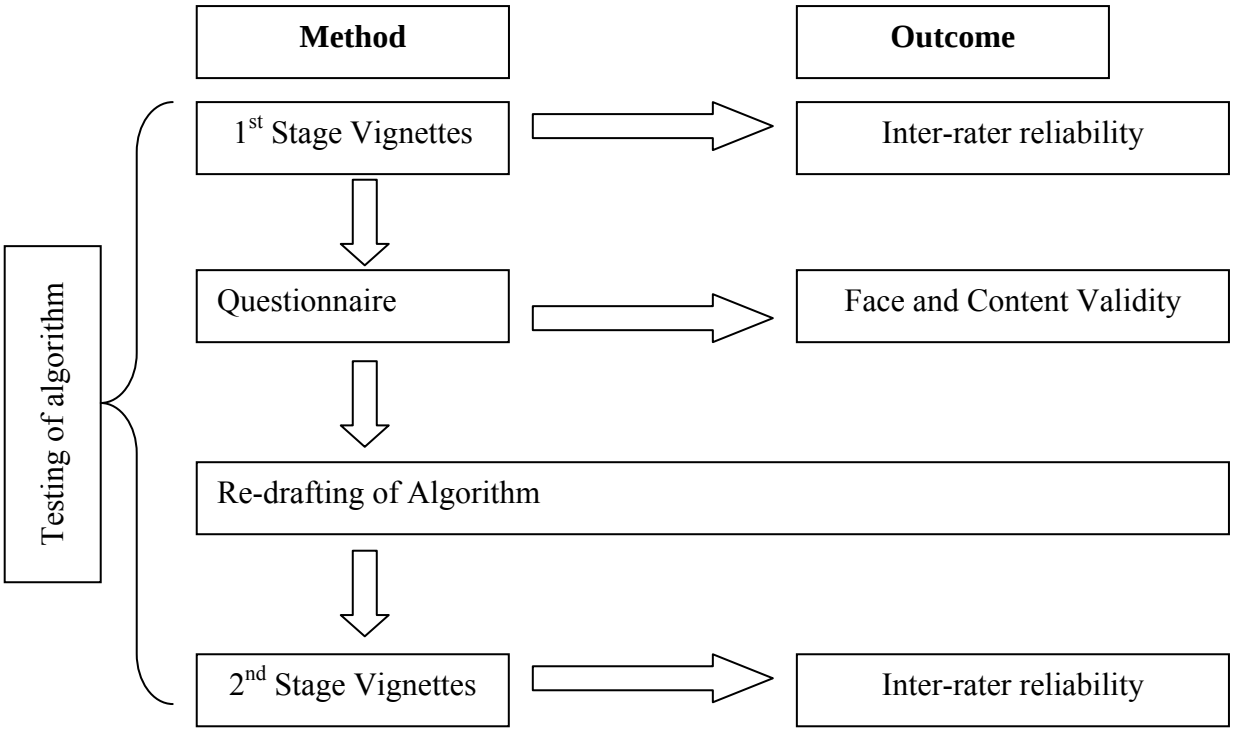
5.5.1. Objectives

The specific objectives were to assess the content of the algorithm, its acceptability to midwives and its impact on consistency of midwives' judgements.

5.6. Methods

Questionnaires and vignettes were used to test the content of the algorithm and its acceptability to midwives (face and content validity) and its effect on consistency of midwives' judgement (inter-rater reliability) respectively. Figure 7 describes the methods used in this stage of the study. Documentation relating to study two is presented in Appendix 2.

Figure 7 Methods used in study two



5.7. Participants and setting

The target population was midwives either currently working in the labour suite or with recent labour suite experience. Two samples of midwives from two

study sites were required. Sample one was used for the first stage which involved vignettes and questionnaires and sample two for the second stage which involved vignettes only.

Sample one was midwives working in a large urban maternity unit in the North of England. This location was chosen to minimise the potential for bias in the subsequent cluster trial, which would be conducted in Scotland. Midwives who had participated in study one were eligible to participate in this study also and there was some overlap between the samples. A seminar was held for midwives at which information about the study was presented. Thereafter, written information, including an invitation to participate and a pre-paid envelope were distributed to midwives to return contact details if they wished to take part. Sample two was midwives who worked in either of two maternity units in Scotland which had agreed to participate in the forthcoming feasibility study (chapter seven). The recruitment process was as described for sample one.

5.8. Methods

5.8.1. Vignettes

A vignette study was conducted to test the consistency of midwives' judgements using the algorithm. This was done in two stages, using midwives from samples one and two. Streiner and Norman (2003) suggest that while both inter-rater and intra-rater reliability may be used, the measurement of intra-rater reliability may be unnecessary. This is because inter-rater reliability

contains the sources of error which contribute to intra-rater reliability as well as sources of error which arise between judges. In this study only inter-rater reliability was assessed.

Vignettes, defined as simulations of real events (Flaskerud 1979), have been widely used in social science, health and decision-making research (Cioffi 1998 1997). They may be developed from a range of sources including literature review, previous research findings and real-life case histories. They are particularly useful in situations where direct observation is inappropriate (Ludwick and Zeller 2001) or where (as in the current study) assessment of the same scenario by a number of judges is required. Vignettes are limited in that they cannot replicate the complex nature of real life (Ludwick and Zeller 2001). Additionally, as the information available is predetermined in the construction of the vignette, there is limited opportunity for the participant to seek further information, as would be possible in a real life situation. Further, participants' responses are hypothetical and must therefore be different to real life judgements (Stolte 1994). Nevertheless, vignettes are useful as a research method because they allow the researcher to control the number and order of cues presented to each participant and because all participants respond to the same stimulus, thereby affording a degree of consistency and control not possible in real life situations (Gould 1996; Hughes and Huby 2002).

A set of forty vignettes were developed by transcribing and anonymising real case histories of women admitted to a maternity unit. The vignettes were

designed to resemble the labour admission page of a woman's case record and contained all the information recorded by midwives on the admission assessment including; a brief summary of the woman's medical and antenatal history, her self reported admission history, the midwives' findings including a general, abdominal and vaginal examination (a sample vignette is shown in Appendix 2). Prior to inclusion, the vignettes were reviewed for face and content validity by six clinical midwives independent of the study.

As the primary purpose of the study was to test the consistency of midwives' judgements using the algorithm, the participants (sample one) were randomly allocated either to receive vignettes and the algorithm (VA1) or to receive the vignettes only (VO) on a 2:1 basis. This design allowed the consistency of midwives' judgements with the algorithm to be compared to that of midwives using clinical judgement alone. Midwives were asked to review the set of vignettes, either using the algorithm (VA1) or clinical judgement alone (VO) and to make the judgement: 'in active labour' or 'not in active labour' for each one.

5.8.2. Questionnaire

Midwives who received the vignettes with the algorithm (VA1) were also asked to evaluate its content, design and acceptability by completing a short questionnaire. Study packs containing study information, the set of vignettes (with or without algorithms and questionnaires) and pre-paid return envelopes were distributed to midwives by post.

Following analysis of all data, the algorithm and vignettes were modified and inter-rater reliability was re-assessed using sample two. The study procedure was as described above, with the exception that all midwives reviewed the vignettes using the algorithm (VA2).

5.9. Ethics

Permission was granted by the appropriate research ethics committees in each area (2002/208; MREC/02/10/21) (Appendix 2). Midwives gave consent to take part in the study having been informed that it was the inter-rater reliability of the algorithm which was being assessed, and not their ability to accurately identify whether or not a woman was in labour. Although the outcome was known to the researchers (as the vignettes were based on real cases), this information was not used to identify the accuracy of midwives' judgements.

5.10. Analysis

Data from the vignettes were entered onto STATA9 and analysed using Cohen's kappa for multiple raters (Fleiss 1971). While a simple calculation of percentage agreement could have been used to measure the agreement between judges, this measure does not take into account the agreement which would be expected by chance alone. In this situation if the midwives chose 'in labour' or 'not in labour' at random then a percentage agreement of 50% could be achieved (Bland 2005). Cohen's kappa is a more reliable measure of agreement as it adjusts the recorded percentage agreement to take account of the agreement which could have been obtained by chance (Altman 1991). The

levels for interpretation of kappa scores are shown on Table 6. For each vignette, midwives were asked to make the judgement 'in labour' or 'not in labour'. However, the midwives could exclude the woman as ineligible for the use of the algorithm (e.g. by judging that she was not a normal, healthy 'prim') or make no decision (missing data). For each stage all data were analysed, including missing data and cases excluded as ineligible, thus presenting the worst case scenario of the inter-rater reliability of the algorithm. Analysis was then performed including only vignettes where complete judgements had been made (i.e. in labour / not in labour). Cohen's kappa analyses for multiple raters were performed for the VA1, VA2 and VO groups providing three kappa scores.

5.11. Findings

5.11.1. Vignettes

Twenty three midwives volunteered to participate in the first stage test of inter-rater reliability (sample one); 16 were sent packs with vignettes and the algorithms (VA1) and seven with vignettes only (VO). Twenty midwives volunteered to participate in the second stage of the study (sample two) and all were sent a study pack with vignettes and algorithms VA2. Table 5 presents the characteristics of the respondents.

Table 5 Characteristics of respondents for study two

Characteristic	1st stage test VO n=7	1st stage test VA1 N=12	2nd stage test VA2 n=17*
Years as a practising midwife			
1-5	3	5	3
6-10	0	4	5
> 10	4	3	8
Clinical Grade**			
E/F	5	8	10
G	1	3	6
H & I	1	1	
Current area of clinical practice			
Delivery suite	4	6	4
LDRP ***	-	-	7
Integrated teams	-	-	5
Postnatal	2	1	
Other	1	5	
Qualifications			
Professional	3	3	10
Degree	2	8	6
Higher degree	2	1	0

*missing data n=1

** Clinical grades: The UK had a clinical grading system for midwives and nurses. Grade G – senior clinical midwife with team leader or ward management responsibility, Grade F –experienced clinical midwife, Grade E/F – junior staff midwife.

*** Labour, delivery, recovery & postnatal rooms

In stage one, twelve (75%) of the VA1 midwives responded, giving a total of 480 possible judgements (i.e. 12 midwives completing 40 vignettes each). All seven VO midwives responded giving a total of 280 possible judgements (i.e. seven midwives completing 40 vignettes each) (Table 6).

Table 6 Inter-rater reliability of midwives' judgements

			All data	Kappa	Complete data	Kappa
1st stage	VO	n=7	280	0.81	278	0.83
	VA1	n=12	480	0.36	459	0.45
2nd stage	VA2	n=17	680	0.68	629	0.86
Kappa Score	Level of agreement		Kappa Score			
<0.20	Poor		0.61 – 0.80	Good		
0.21 – 0.40	Fair		0.81 – 1.00	Very good		
0.41 – 0.60	Moderate					

When all data were included, inter-rater agreement for VA1 midwives (n=12) was only fair (K=0.36). There were 21 missing or incomplete judgements out of the possible total of 480, when these were excluded the Kappa score was 0.45 which still represents only moderate agreement. Fifteen of the missing judgements were accounted for by three midwives who reported that they were unable to classify between four and six cases each due to lack of information in the vignettes. Inter-rater agreement for VO midwives (n=7) was very good (K= 0.81). There were only two incomplete judgements out of a total possible 280, when these were excluded the Kappa score was 0.83.

5.11.2. Questionnaire, redrafting of the algorithm and amendment of vignettes

All twelve VO midwives also returned the evaluation questionnaire for the algorithm. Eleven (92%) reported that the algorithm was either easy or very easy to complete, however most identified some snags either with the algorithm or the vignettes and were able to make suggestions for their improvement (Table 7).

Table 7 Content, design and acceptability of the algorithm

Question	VA1 n =12		
How easy to complete was the algorithm			
Easy/very easy	11		
Not easy/difficult	1		
	Yes	No	Specific comments
Did you experience any snags in completing the algorithm?	10	2	Vignettes contained insufficient information on pain (7). Use of term "primigravid" excludes women who have had miscarriage (6), algorithm layout problem (3)
Was anything missed out?	5	7	Definition of normal labour (1), assessment of the woman's emotional state (1), guidance on management of women not in labour (3)
Should anything be removed?	8	4	Assessment of contractions should be less prescriptive (8)
Additional suggestions or comments	10	2	Layout improvement (4), Useful tool for midwives and women (3), not useful (1) Midwives will prefer clinical judgement (1), Contractions and cervical dilatation should be weighted above other cues (2)

The presence of moderate abdominal pain was central to the algorithm however, in about one third of vignettes pain was not explicitly described. Although the vignettes were transcripts of real histories, the women were often described as 'coping well' or 'distressed'. This resulted in midwives being unable to complete the algorithm. Both the algorithm and the vignettes were modified, with reference to the midwives' comments. For example, an appropriate description of pain was added in the vignettes and the term 'primigravid' was replaced by the correct term 'primiparous' in the algorithm.

5.11.3. Re-testing the algorithm

In stage two twenty midwives initially volunteered to participate and seventeen midwives (85%) responded, giving a total of 680 possible judgements (i.e. 17 midwives completing 40 vignettes each). The kappa score when all data were included was 0.68 indicating good inter-rater agreement (Table 6). However, this included a number of missing or incomplete judgements (51 out of a possible 680). The majority of these were accounted for by three midwives who consistently made the error of confusing the terms primigravid and primiparous, each excluding at least 11 out of their set of 40 vignettes of women who had parity 0+1. When incomplete judgements were excluded the kappa score was 0.86, indicating a very good level of agreement. The final version of the algorithm is presented in Appendix 3.

5.12. Discussion

Testing the content and design of the algorithm highlighted a number of weaknesses, in particular with regard to the terminology used. The term primigravid was initially used, but midwives (VA1) in the first stage of the study, identified that the correct term was primiparous (the intention was to include all women giving birth for the first time), and the algorithm was consequently changed. Subsequently, a number of midwives in the retest (VA2) erroneously excluded a group of cases which they judged were ineligible (not ‘prims’). This demonstrates the confusion which can arise over simple terminology and highlights the importance of testing prior to implementation of any clinical decision support tool.

In this study the consistency of midwives' judgements was explored rather than their accuracy. Consistency of judgement is important, as inconsistent judgements must be inaccurate; at least some of the time, therefore testing the algorithm for consistency is an essential step in assessing the potential usefulness of the tool. However, consistency is not a guarantee of judgement accuracy as clinicians (aided or un-aided) may make consistent but wrong judgements.

Vignettes have been used in social research for 20 years (Flaskerud 1979), their strength is that the same cues are presented in the same order to every participant. However there are limitations to this method, in particular, vignettes are only as good as the information they contain. In the first test of inter-rater reliability inadequate information on pain was included, despite the fact that the vignettes were transcribed from real case histories and reviewed by a group of clinicians. This meant that midwives were unable to complete their judgements for a number of cases. It is interesting that the case histories (written by midwives) from which the vignettes were transcribed did not record an assessment of pain, but instead included euphemisms such as 'distressed' or 'coping well.' Studies on labour pain have identified that midwives consistently under-rate the intensity of pain experienced by women (Niven 1993). Use of this algorithm prompts the midwife to make an objective assessment of the pain experienced by the woman and this may confer some benefit, however, further research is required in this area.

A further limitation of vignettes is that they cannot replicate the uncertainty of clinical judgement in the real world. In this study midwives who used clinical judgement alone (VO) demonstrated highly consistent judgements (this consistency was ultimately only matched by the algorithm-aided judgements in the second stage), this begs the question, why is decision support required? However, there is ample evidence of the difficulty which is experienced in making this judgement in a real world setting (Ball and Washbrook 1996; Hemminki and Simukka 1986; Thornton and Lilford 1994; Holmes et al. 2001; Klein et al. 2003). In addition, one study has reported that the application of strict criteria to diagnosis of labour did have an impact on clinical outcomes by reducing oxytocin use (McNiven et al. 1998).

Overall the results of this study demonstrated that the algorithm did comprise the key cues for diagnosis of labour, that these were presented in a form which was acceptable to midwives and the algorithm did achieve a high level of consistency of judgement between midwives. However, the usefulness of the algorithm now required to be tested in a clinical trial to determine whether its application in a real world setting would result in improved diagnosis of labour as evidenced by improved clinical outcomes for women.

CHAPTER 6: DEVELOPING THE METHODS FOR A CLUSTER RANDOMISED TRIAL TO INVESTIGATE THE USE OF AN ALGORITHM FOR THE DIAGNOSIS OF ACTIVE LABOUR IN PRIMIPAROUS WOMEN

6.1. Introduction

In the preceding chapters the problem of diagnosis of labour was highlighted and an algorithm which aimed to improve the diagnosis of labour, was developed using both qualitative and quantitative methods. Initial testing of the algorithm demonstrated that it had good face and content validity as well as a very good level of inter-rater reliability. However, in order to determine whether it was effective in improving diagnosis of labour a clinical trial was required. This chapter describes the development methods for study four, the CRT (Documentation relating to study four is presented in Appendix 5). Following development, these methods were tested in a feasibility study (chapter seven).

6.2. Aim

The aim of the CRT was to compare the effectiveness of the algorithm for diagnosis of active labour, in healthy primiparous women, with standard care in terms of maternal and neonatal outcomes.

6.3. Objectives

The objectives were to determine the effectiveness of the algorithm in terms of a reduction in the use of oxytocin for augmentation of labour, rates of medical

interventions in labour, rates of instrumental and operative delivery and in the admission of women subsequently found not to be in labour.

6.4. Study design

Documentation relating to study four is presented in Appendix 5. In a CRT groups of participants, frequently existing social units, are randomly allocated to different treatments. Allocation is by group, rather than by individual, as would be the case in a standard randomised controlled trial. In trials where randomisation to experimental and control group takes place at the level of the individual, outcomes for each participant are assumed to be independent and data analysis is conducted at the level of the individual thus maximising the power of the study (Donner 1998; MRC 2002). However, there are specific circumstances in which it is not appropriate to randomise at the level of the individual participant and a larger unit (cluster) may be used (Torgerson 2001).

In some cases it may not be possible for the intervention to be administered at an individual level. For example, in studies involving treatment of water supplies it would be difficult for the individual to avoid the experimental treatment (Luby et al. 2006). Another case would be where the intervention may act at both a community and individual level; for example, studies into the effects of programmes of vaccination. In a study to determine whether vaccination of care home staff against influenza indirectly protected residents (Hayward et al. 2006), vaccinations were offered to all staff in the intervention

group care homes, and potential health benefits were observed from the care home residents.

In other studies, while it may be possible to deliver the intervention to the individual there may be a risk of between-participant contamination. In studies of clinical guidelines or educational packages it would be difficult for the professional to limit the intervention to specific individuals in their care and therefore there would be a risk of contamination between study groups. In a study of postnatal care (MacArthur et al. 2002), the intervention comprised an innovative package of care, including midwife education and postnatal symptom checklists used by midwives. As midwives could not selectively apply their knowledge to individually randomised participants, and because more than one midwife could provide care for any particular woman, contamination between study groups would have occurred if randomisation at the level of the individual was used. For that reason, a cluster design was implemented in which the unit of randomisation was general practitioner group practices.

A CRT was chosen in the current trial because the intervention, the algorithm, was aimed at the clinical practice of midwives. Individual randomisation of women could not be used because midwives could not be expected to provide two types of care, using the algorithm for some women and clinical judgement alone for others, without contamination between groups. Midwives could not be randomly assigned because of the possibility that study materials would be passed between individuals in different study groups and, as in the postnatal

care study (MacArthur et al. 2002), more than one midwife could be involved in the care of an individual woman.

Use of a CRT design has implications at each stage of study development and implementation, these will be discussed below.

6.5. Statistical power

Although the randomisation of groups rather than individuals may be essential to avoid contamination between study groups, it leads to a significant loss of statistical efficiency in the trial (Donner 1998; Bland 1997). This occurs because individuals within a cluster tend to have characteristics in common and will be more similar to each other than to individuals in another cluster (Donner and Klar 2000; Bland 1997; MRC 2002). There may be a number of reasons for this. Individuals may have chosen the cluster to which they belong, for example, by choosing to live in a particular area or register their children at a particular school. Patients attending their local hospital or GP practice are likely to live within the catchment area and to have similar demographic characteristics. Additionally, they are likely to receive similar types of care because of local policies or guidelines or because caregivers within a particular hospital are likely to share a similar philosophy of care. This means that individuals within a cluster cannot be assumed to be independent from each other and are more likely to have similar clinical outcomes than individuals from a different cluster. This is known as the intracluster correlation and is quantified by using the intracluster correlation coefficient (ICC) which takes a value of

between 0 and 1, with a higher ICC representing greater similarity within a group (Donner and Klar 2000; Campbell et al. 2000). The intracluster correlation has a major effect on both the required sample size and the analysis of the study.

Standard statistical methods assume that the unit of randomisation and analysis are the same and that the individual participant and individual observations are independent (Mollison et al. 2000). Although, the unit of randomisation in a CRT is the cluster, the unit of analysis is generally the individual within that cluster. If standard statistical methods are applied to a CRT they result in a reduction in the standard error and p values, which leads to an overestimation of the effect of the intervention (Mollison et al. 2000; Donner and Klar 2000). This problem is increased where the ICC is high (the individuals within the cluster are more similar) and cluster sizes are large (Bland 1997; Donner 1998). The within and between cluster variation must therefore be assessed and taken into consideration in calculating the study sample size.

Calculation of the sample size requires an assessment of the number of potential clusters available, the potential size of clusters and a prior assessment of the ICC. This requires an analysis of data on the distribution of primary outcome measures within the study population, obtained from a review of existing data sources or more accurately from pilot study data. The standard sample size calculations for independent observations require consideration of the clinically relevant potential effect attributable to the study intervention

(Mollison et al. 2000). In a CRT the sample size must be multiplied by the variance inflation factor otherwise known as the design effect. This is the ratio of the number of participants required in a cluster study to the number required in an equivalent individually randomised trial. The design effect measures the magnitude of the effect of clustering, it takes into account the size of the clusters and the ICC, and provides the calculation for confidence intervals and test statistics. For equal sized clusters the design effect is given by the equation: 1+ (m-1) x ICC where m is the cluster size. Greater gain in statistical power is achieved by increasing the number of clusters rather than by increasing the number of individuals within the clusters, because increasing the cluster size increases the design effect. This may render some trials impractical because of the lack of available clusters.

6.6. Participants

While standard RCTs have one level of participation, the individual, CRTs are more complex having at least two, and often three levels (Hutton 2001; MRC 2002). The number of levels is dependent on the nature of the intervention. The first level is that of the cluster. Typically the cluster will be a social unit, for example a family, a village, or other pre-existing group such as patients belonging to a general practitioner group practice (Donner 1998). This is the unit of randomisation, or allocation to study group. The second level is that of the experimental unit, the participants at whom the study intervention is targeted. These may be the health care professionals who work within the cluster (as in a study of clinical guidelines) or the individual patient receiving care from the cluster. The third level is the unit of observation or inference. It is

at this level that study outcomes are measured. For example, in a study of an educational package for perineal suturing, the maternity unit could be the cluster, midwives receiving the educational package might be the experimental level and the perineal healing of individual women could be the observational level. In some CRTs however, there may only be two levels since the cluster level may also be the experimental level. An example would be where a whole village is allocated to receive vitamin supplements (the village being both the cluster and experimental levels) and health outcomes are measured at the level of individual inhabitants (the observational level).

There were three levels of participation in the current trial. These were:

Cluster level. Trial entry and randomisation took place at the level of the maternity unit. The Expert Group on Acute Maternity Services (Scottish Executive 2002) defined levels of maternity care in Scotland according to specific criteria relating to services, staffing and birth rate. Maternity units defined as IIb, IIc and III (Table 8) within the trial period were eligible to participate. These units had the facilities to provide oxytocin for augmentation of labour, while smaller units had to transfer women requiring oxytocin to a referral centre. As oxytocin for augmentation of labour was the primary outcome measure, IIb units were considered to be the minimum eligible for participation. Reorganisation of maternity services in Scotland meant that the number of maternity units within each of these categories declined during the course of the studies presented in this thesis. During the trial development period there were 20 maternity units classed as being IIb- III.

Table 8 Levels of maternity care in Scotland (EGAMS, 2002)

Level of care	Location of delivery	Lead carer	Clinical situation	Annual births
Ia	Home	Midwife	Normal pregnancy and labour	
Ib	Stand-alone community maternity unit	Midwife (GP)	Normal pregnancy and labour	
Ic	Community maternity unit adjacent to non-obstetric hospital	Midwife (GP)	Normal pregnancy and labour	
Id	Community maternity unit adjacent to maternity unit	Midwife (GP)	Normal pregnancy and labour	
IIa	Consultant-led maternity unit with no neonatal facility	Consultant Obstetrician(plus midwife)	Low risk pregnancy and labour	<1,000
IIb	Consultant –led maternity unit with on-site neonatal facility	Consultant Obstetrician plus Midwife	Low to medium risk pregnancy and labour	<1,000
IIc	Consultant –led maternity unit	Consultant Obstetrician plus Midwife	Low and most high risk pregnancies and labour	1,000 – ≥3,000
III	Consultant –led specialist maternity unit	Consultant Specialist in maternal fetal medicine, Midwives and others	Complex and high risk pregnancies and labour	>3,000

Experimental level. This level was present in maternity units in the experimental group only. The intervention in the trial (the algorithm) was targeted at the clinical practice of midwives; therefore midwives using the algorithm were study participants. All midwives who regularly admitted women to labour suites, within the maternity units allocated to the experimental group, were eligible to participate.

Observational level. This level was present in both experimental and control groups. The outcomes of the trial were assessed from those intended to benefit from the intervention, in this case healthy primiparous women. Women were eligible for the trial if they were; primiparous, presenting for admission in spontaneous labour, at term and assessed as low risk based on criteria used in

previous intrapartum care trials (Hundley et al. 1994; MDU 1995; Cheyne et al. 2003). Eligibility criteria are shown in Appendix 5. In order to reduce confounding variables multiparous women were excluded. Although the principles of diagnosis of labour are the same for both primiparous and multiparous women, there are significant differences in the way in which their labour would be expected to progress (O'Driscoll et al. 1973). A woman's previous experience of labour may also influence her current intrapartum care.

6.7. Allocation to study group

In a CRT allocation to study group takes place at the level of the cluster. As with all controlled trial designs randomisation is used to ensure that factors which influence study outcome are equally distributed between groups (Treasure and MacRae 1998). There are several ways by which random allocation may be approached; unrestricted allocation, restricted allocation and minimisation (Donner 1998; Treasure and MacRae 1998; MRC 2002; Altman and Bland 2005). Using an unrestricted method, allocation to group is completely randomised with no stratification or matching. This method is appropriate in studies in which a large sample is available. However, since the unit of study allocation is the cluster the actual sample available at the level of allocation is reduced, increasing the likelihood of differences arising between study groups. In studies where restricted allocation is used, pre-identified baseline characteristics, which are likely to correlate with study outcomes, are used to divide clusters into strata or matched pairs. Clusters within strata or pairs are then randomly allocated to study group. Restricted allocation increases the likelihood of balance between study groups; however the

resultant gain in statistical power may be lost because of loss of information on between- cluster variability (there is a reduction in degrees of freedom available for estimation of error) and it may therefore be difficult to estimate the ICC (MRC 2002). Restricted allocation through stratified and matched randomisation designs are not appropriate where the available sample size is small (Altman and Bland 2005).

Minimisation is a method which can be used in studies with a small sample size. Using minimisation allocation to study group is not based on random allocation alone. The first unit to enter the study is randomly allocated to group, thereafter clusters are purposively allocated in order to maximise balance between groups for pre-identified characteristics which may predict the outcome (Treasure and MacRae 1998; Altman and Bland 2005). In the current trial minimisation was used as the means of group allocation. Presence or absence of an on-site midwife managed birth unit (MBU) was the balancing variable; chosen because midwives providing care within a MBU would be expected to share a similar philosophy of care in aiming to provide low intervention care for normal healthy women. This would be anticipated to correlate with a lower use of oxytocin and medical intervention in labour. Where minimisation is used it is essential that allocation to group is performed by an independent person, as the method is not as unpredictable as randomisation. In the current trial, group allocation was performed by the statistician (Martin Bland), who was not involved in recruitment. Blinding of participants (maternity units and midwives) or data collectors, to study group

was not possible in this trial because use of the algorithm was evident in maternity units in the experimental group and in the case records of women who participated in the trial.

6.8. Study groups

6.8.1. Intervention:

The intervention in this trial was the use of the algorithm for diagnosis of active labour in primiparous women described in chapter five. Midwives were asked to use the algorithm during their assessment of women on admission, to assist them to diagnose active labour, recording their judgement on the algorithm. The algorithm was printed on duplicate paper, once completed one part was retained in the woman's case record while the other was collected by the local study co-ordinator.

6.8.2. Control:

Eligible women admitted to maternity units allocated to the control group received standard care for their particular unit. Standard care in relation to admission of women in labour varied between maternity units. However, a telephone survey (chapter four) of maternity units in Scotland found that none had written guidelines for the diagnosis of labour.

Following the admission assessment, women in both arms of the trial received standard care for their maternity unit.

6.9. Study entry and consent

The multilevel nature of CRT designs has implications for study ethics (Edwards et al. 1999; MRC 2002; Hutton 2001). While all normal research ethics principles apply to a CRT, there are two ethical principles which may present specific problems. These are the principle of voluntary consent to participation and the freedom to withdraw at any time during the course of the experiment. Because the cluster is the unit of entry and allocation in a CRT, it may be difficult for an individual within a cluster to withhold consent to study entry or to withdraw during the course of the study. In relation to consent CRTs may be divided into two types (MRC 2002). In the first type the intervention is delivered at cluster level and there is no opportunity for an individual to choose to participate or to withdraw; for example, studies involving medication of water supplies. In the second type while the cluster remains the unit of study entry and group allocation the intervention is targeted at an individual level and it would be possible for an individual to choose whether or not to accept the intervention. For example, in a CRT of a weight reduction intervention comprising hypnotherapy, GP practices could be allocated to experimental and control groups. However, individual patients would then be able to accept or decline the study intervention. Ideally consent should be sought at each level of participation.

In all CRTs consent for study entry and randomisation must be obtained from a gatekeeper or series of gatekeepers, independent of the research team, who have the authority to act on behalf of the cluster, and must act in the interest of

the cluster (MRC 2002). In trials where a distinct experimental level exists, for example, a study of guideline implementation, or educational package, individuals should be asked for consent to accept the intervention. At the observational level it may not be possible for individuals to choose whether to receive the intervention or not. In the case of studies of guidelines, the intervention will direct the practitioner to work or think in a particular way and they may be unable to alter this on a patient by patient basis. In this situation consent may be gained or withheld for data collection and administration of follow–up questionnaires only. However, in some situations where only anonymised data are used, individual consent may not be necessary. At the experimental and observational levels in a CRT, consent is being sought following randomisation. This is similar to Zelen's randomised consent method (Donner 1998; Hutton 2001). Selection bias may arise in the situation where cluster consent has been gained but a substantial number of individuals subsequently withhold consent, either to accept the intervention or to permit collection of data (Torgerson 2001). Bias occurs because refusal to participate may be more or less likely in particular groups and will not be randomly distributed. Some degree of selection bias is likely in most CRTs.

Ethical approval for the current trial was granted by the Multicentre Research Ethics Committee for Scotland (05/MRE10/31) (Appendix 5). A Principal Investigator (PI) was appointed for each participating unit. In most cases this was the Head of Midwifery or another senior midwife. Site specific ethical approval was granted in each area. The study developed a strategy for

negotiation of access and consent gaining based on one which was described by Walker et al. (2000). The strategy was as follows.

Cluster level: The Heads of Midwifery in the eligible maternity units were initially approached, and an individual unit plan for information giving and consent for trial entry was devised with them. This involved discussion with local stakeholders including senior midwives, the Clinical Director and in some cases lay representatives. Following this discussion the Clinical Director was asked to give consent to trial entry and group allocation, on behalf of the unit. Thereafter, a clinical midwife based in the delivery suite of each unit was nominated to be the local trial co-ordinator and an individual plan was made for recruitment and consent of midwives.

Experimental level: This level was present in maternity units in the experimental group only. All midwives who admitted women in labour were fully informed about the trial. Training workshops and individual contacts were provided for each midwife. Each was given a workbook containing study information, algorithm and vignettes of case histories. Workshops included discussion about the need for evidence based practice and the method and purpose of the CRT, as well as information on decision-making and the effect of decision support. Midwives were provided with the information they required to use the algorithm and complete the trial documentation, in particular seeking consent from women in early labour. Midwives were then asked to give written consent for participation. This process took approximately one month in each unit,

although a rolling programme of information giving and consent was required in some units to accommodate staff rotation programmes and team models of care. Only the midwives in the intervention sites received this training, minimum contact was made with midwives in control units to reduce the potential of a Hawthorne effect (Braunholtz et al. 2001).

Observational level: Studies of intrapartum interventions present additional ethical challenges, in particular in relation to consent, study entry and randomisation of participants who are in a situation of stress and vulnerability. Hundley and Cheyne (2004) reviewed RCTs of intrapartum interventions and identified particular issues in relation to intrapartum studies. These issues included: the fact that eligible women cannot be reliably identified until labour admission; selection bias is inevitable where caregivers are responsible for recruitment and will use personal clinical judgement in deciding whether it is appropriate to approach a particular woman; and pre or post randomisation losses will occur, the magnitude of each depending on the timing of randomisation. Three main methods of recruitment and consent gaining were identified, antenatal recruitment and randomisation, recruitment and randomisation on admission in labour and staged recruitment and randomisation. Each method had particular strengths and weaknesses; however, no method eliminated the problems identified above.

In the current trial data for the main trial outcomes could have been collected using anonymised data where no individual consent would have been required. However, a health economics evaluation was planned in parallel with the CRT

and involving a subset of women. Therefore, it was necessary to seek consent for data collection and administration of questionnaires. A two stage procedure was adopted. Primiparous women in each unit were given information about the study at their antenatal clinic visit between 34-36 weeks gestation. In the units allocated to the experimental group women were checked for eligibility (by the admitting midwife) when they sought admission to the labour suite. Eligible women were then given a full explanation of the trial, including written information, and were asked to give consent. Midwives then used the algorithm to support them in making their judgement as to whether or not the woman was in active labour. In the control group units, women were asked for consent during their postnatal hospital stay to reduce the potential for a Hawthorne effect (Braunholtz et al. 2001) which would have occurred if midwives in labour suites were asked to seek consent.

6.9.1. Study information

The notion that merely the knowledge that one is participating in a research study will alter the normal behaviour of participants has been widely discussed (Wickstrom and Bendix 2000; Braunholtz et al. 2001; McCarney et al. 2007). One of the reasons for conducting a CRT is to reduce contamination between study groups however, in order to ethically conduct a trial it is necessary to make information available to participants. In addition, information may be gathered by those either directly or indirectly involved in a trial and this may result in a change in behaviour. In the current trial information was provided to those involved directly and indirectly as follows:

All Supervisors of Midwives in Scotland were given a briefing sheet about the trial (Appendix 5), not including an example of the algorithm, prior to its implementation. This was considered essential because of the role of the supervisor in supporting midwifery practice and protection of the public.

At cluster level Heads of Midwifery and Clinical Directors of all participating units (experimental and control) were given full information about the nature and purpose of the trial, prior to randomisation. This did not include access to the algorithm which was only made available to experimental units following allocation to group.

Within the units allocated to the experimental groups there was no restriction of information about the study. Participating midwives (those working in labour suite) were given full information about the trial, including algorithms as described above. Midwives indirectly involved for example, midwives working in the community, antenatal clinics and postnatal areas, were given summary trial information (Appendix 5), however, full information was available if requested. Women who were potentially eligible for trial participation were given a full explanation of the purpose of the trial. Posters providing information about the trial were displayed in antenatal clinics and labour suites.

Within units who were allocated to the control group information about the trial was restricted, in particular, no access was given to the algorithm. Information

was only provided to midwives who were indirectly involved in the study by providing trial information to women, for example, midwives working in community and antenatal clinics and postnatal ward areas. This explained that a national study was being conducted which aimed to improve diagnosis of labour. No information was provided to midwives working in labour suites. Local trial co-ordinators were not given access to the algorithm. Women were provided with information about the study during the antenatal and postnatal periods.

6.9.2. Outcomes measures

The review of decision-making literature (chapter three), identified that most RCTs of decision support tools have used measures of the quality of the process of care, while relatively few have used indicators of improvement in clinical outcome. This parallels the coherence or correspondence debate in studies of human judgement and decision-making (Hammond 1996); that is, whether a good judgement is one that results in a good outcome (correspondence theory) or one in which the judgement itself follows the rules of rationality (coherence theory). As discussed in chapter five, for the algorithm to be considered useful in a clinically relevant sense, it would have to produce evidence of improved clinical outcomes for women. Therefore, in this trial clinically relevant outcomes, rather than measures of the decision process, were chosen to determine the efficacy of the algorithm. The choice of measures is discussed below.

6.9.3. Sensitivity and specificity

An appropriate measure of the efficacy of a diagnostic test would be its sensitivity and specificity. Sensitivity relates to the proportion of instances in which a positive test result corresponds with the presence of the condition, a true positive. Specificity relates to the proportion of instances where a negative test result corresponds with the absence of the condition, a true negative (Altman and Bland 1994). Although the algorithm was a diagnostic aid rather than a test it is possible to consider whether sensitivity and specificity would apply. In this situation a true positive result (sensitivity) would be a case where the use of the algorithm correctly indicated that a woman was in active labour and a true negative (specificity) would be a case where the algorithm correctly indicated that a woman was not in labour. These outcomes appear to be clear cut, however, they were not used as outcomes in the current trial for the following reasons. Following assessment using the algorithm, a period of time during which no clinical intervention took place, would be required in order to determine whether a true positive or negative had been obtained. Friedman (1989), suggested that during the active phase of labour a woman's cervix would be expected to dilate at a predictable rate of one centimetre per hour. However, the appropriateness of ridged adherence to time parameters has been questioned (Walsh et al. 2004). Moreover, it is accepted that the latent phase is a poorly-defined period which may extend up to 20 hours without detriment (Austin and Calderon 1999), therefore, any time period applied in this trial would be arbitrary. Furthermore, in a real world situation, it would clearly be both impractical and unethical to limit the care provided to women in order to 'await events' for the purpose of research. Where a woman is diagnosed as

being in active labour and subsequently is judged not to have made the anticipated amount of labour progress, she is likely to be diagnosed as having dystocia or failure to progress and receive augmentation of labour. In this situation it would not be clear whether this was a false positive result or a true positive case who received medical intervention too soon. Conversely, a woman who is diagnosed as not in active labour may opt to remain in hospital and may then receive oxytocin for labour augmentation; again, it would not be clear whether this was a false or true negative. For these reasons sensitivity and specificity were not considered suitable as primary outcomes in this trial; however data relating to admission of women in labour, as well as discharge of women following labour assessment, were collected.

6.9.4. Caesarean section

The high and rising rate of caesarean section is a concern for the maternity services in the UK (Thomas and Paranjothy 2001). One of the two most frequent reasons for performing a caesarean section is a diagnosis of failure to progress in labour (Thomas and Paranjothy 2001; SPCERH 2003; ISD 2006) and therefore it is clearly an important and relevant clinical outcome in relation to the trial. The overall rate of caesarean section is around 21-24% (Thomas and Paranjothy 2001), therefore it is possible that caesarean section should have been the primary trial outcome. However, when elective caesarean section is excluded, the rate of emergency caesarean section is around 11% (SPCERH 2003; ISD 2006) and consequently a very large trial sample would be required. Furthermore, failure to progress is only one of the reasons for an emergency caesarean section and it would be necessary to differentiate

between that and other reasons. A woman who has a caesarean section due to failure to progress will already have had a number of medical interventions in her labour; primary among these is the use of oxytocin. Therefore, use of oxytocin and other medical interventions in labour were considered to be the most appropriate outcomes in the current trial.

6.9.5. Oxytocin for augmentation of labour

Oxytocin use was chosen as the primary study outcome because it is the principal treatment (and key marker) of slow progress in labour (NICE 2004). When a woman is admitted while not in labour, or while in the latent phase, it is suggested that the 'clock starts ticking' (Simonda, 2002) and she is likely to be diagnosed as having slow progress or failure to progress in labour. In these circumstances oxytocin is the principal treatment and is an objective marker of a labour which is considered dysfunctional. Administration of oxytocin in itself requires a woman to have intensive monitoring of labour, including continuous electronic fetal monitoring (EFM) and she is more likely to require epidural analgesia and instrumental or operative delivery (RCOG 2001) as discussed in chapter two. Reduction in the use of oxytocin for primiparous women in spontaneous labour would represent a significant improvement in clinical management. In addition, a previous study of the use of explicit criteria for admission in labour (McNiven et al. 1998), identified a reduction in the use of oxytocin in labour. Oxytocin use is a reliably documented, clinically important intervention; this made it an appropriate primary trial outcome.

The trial outcomes chosen were:

Primary Outcome:

- Use of oxytocin (any dose) for augmentation of labour.

Secondary Outcomes:

- Artificial rupture of membranes
- Vaginal examination
- Use of analgesia including epidural
- Admission management; number of admissions prior to labour, time spent in labour ward, time in active labour
- Mode of delivery
- Intrapartum complications
- Neonatal outcome (APGAR score, neonatal resuscitation and admission to the neonatal unit (NNU).
- Unplanned out of hospital births (Born Before Arrival BBA).

6.10. Data collecting and monitoring

Data collection tools were developed for this trial or adapted from previous intrapartum RCT (Hundley et al. 1994). Where several copies were required paperwork was printed on duplicate (or triplicate) paper in order to minimise

work required by clinical midwives (Appendix 5). Documentation was tested during the feasibility study and subsequently changed if necessary.

Trial implementation was staged, with a planned data collection period of ten months in each unit. With the exception of women's consent forms and algorithms, unit level clinical data collection and trial paperwork were completed by the local trial co-ordinators. A secretary based at the Nursing Midwifery and Allied Health Professionals Research Unit performed all data entry into an Access database.

It has been suggested that every trial should incorporate some form of data monitoring (Grant el al. 2005; Williams 2006). This may include monitoring the conduct of the trial as well as the quality of the data collected and may involve the formation of a data monitoring committee. The remit of the data monitoring committee may vary between trials but central to their role is the monitoring of data in relation to safety and benefit. A data monitoring committee should be multidisciplinary and independent of the study, having the authority in some circumstances to stop a trial prematurely.

In the current trial the author of this thesis (HC) firstly monitored the early returns for each unit, with the aim of ensuring that recruitment rates and compliance with study protocols were meeting agreed milestones. This allowed rapid intervention in units which were experiencing difficulty in implementing the

trial and ensured that progress was satisfactory. In order to ensure the quality of the data collected and entered, a minimum of five data forms from each centre, up to a maximum of 10% of the total data forms were audited for completeness and accuracy of data entry. Data forms were initially checked for completeness by the trial secretary, incomplete forms were returned to the appropriate unit for amendment. Accuracy of data entry was checked by comparing data base entry with the data collection forms for 10% of the study sample. A data monitoring form was completed noting any errors which were then corrected. A proportion of data forms (10%) were audited revealing 88% accuracy of data entry, with 12% of cases having one data entry error which was corrected.

A data monitoring committee was formed. The committee was independent of the study and multidisciplinary comprising a consultant obstetrician, consultant paediatrician, consultant midwife and chaired by a medical statistician. The group met and agreed its operating procedures, namely that they would immediately be informed of the occurrence of severe adverse events in the intervention group and that they would review the occurrence of all severe adverse events (in experimental and control groups) at the mid-point of the study. Severe adverse events were defined as maternal or neonatal death.

6.11. Clinical governance

All study documents (consent forms, data collection forms, questionnaires and audio -tapes) were stored in a locked metal filing cabinet in the research office.

Computer records were stored on a computer designated for the study and password protected. All participants were allocated a unique study number which was used to link data collection instruments. Only anonymised data were recorded on data collection forms. The data will be securely archived for ten years and will then be destroyed in accordance with the University of Stirling's procedures.

6.12. Preliminary sample size calculation

The statistical power calculation in the current trial was appropriate for an unmatched CRT design. Information was required on the ICC and the distribution of the number of deliveries by maternity unit. No data on ICCs for oxytocin use were available through routinely collected data, therefore a feasibility study was required to collect data specifically for the CRT. A preliminary sample size calculation was initially performed as follows, and this was revised using data collected during the feasibility study.

Data routinely collected by the Information and Statistics Division (ISD) of the National Health Service in Scotland (1999) on rates of normal and instrumental deliveries in Scotland were used. These data allowed the ICC for emergency caesarean sections across maternity units to be estimated as 0.005. This ICC for caesarean section was then extrapolated to the use of oxytocin to enable the preliminary sample size to be determined. Due to the extrapolation of the ICC and the imprecision in its estimation, the sensitivity of the sample size to variability in ICC required to be further explored in the feasibility study. It was

initially estimated that to detect a difference of 10% in the proportion of women receiving oxytocin for augmentation of labour, from 40% to 30%, with 80% power and assuming an ICC of 0.005, a total of eight maternity units, with an average cluster size of 200 would be required. Assuming an ICC of 0.01, a total of 12 maternity units, with an average cluster size of 200 would be necessary to detect a difference of 10%. A 10% reduction in the proportion of women receiving oxytocin was the level judged to be clinically relevant by the trial research team. It was clear from this preliminary calculation that a feasibility study was required both to provide accurate data for the sample size calculation and to determine whether a sufficient number of maternity units within Scotland would be available to participate and so make a CRT feasible.

6.13. Analysis

As with sample size calculations, data analysis of cluster trials must take into consideration the effect of clustering. There is no optimum method of analysis for cluster trials (Mollison et al. 2000) and the choice of methods will be affected by factors such as: the unit of inference i.e. whether the study outcomes of interest are measured at cluster or individual level; the number of available clusters; the size of clusters; and the variability of cluster size. There are two main approaches to analysis: analysis at cluster level and individual level analysis. Where analysis is conducted at cluster level a summary measure for each outcome is calculated for each cluster, for example, the cluster mean. This overcomes the problem of non-independence of data, providing one data point for each cluster which may be analysed using standard statistical tests (Kerry and Bland 1998; Mollison et al. 2000). Alternatively, where sufficient

clusters per group (at least ten) are available standard statistical tests (t-tests or Chi square) may be adapted to account for clustering effects and then applied to individual level data. More advanced statistical techniques, which take account of the hierarchical nature (individual and cluster levels), may be used. For example multilevel, hierarchical regression modelling accounts for clustering and permits individual and group characteristics to be included (MRC 2002). The specific analysis conducted in the current trial is described following the revision of the sample size in the next chapter.

6.14. Summary

This chapter has described the development of the methods for the CRT. The MRC framework (MRC 2000), highlights the importance of conducting an exploratory trial. At this stage key components of the trial methods may be tested and modified, if necessary prior to implementation of the full trial. An exploratory trial may be used to establish the feasibility of a full trial as well as to collect data necessary for calculation of the power of a large trial. The following chapter describes a feasibility study (study three) that was conducted prior to the implementation of the CRT.

CHAPTER 7: STUDY THREE: FEASIBILITY STUDY

7.1. Introduction

Previous chapters have described the development of the algorithm and its preliminary testing. Its efficacy in improving clinical outcomes for women next required to be tested using a CRT. Development of the CRT methods was described in chapter six. This chapter describes the feasibility stage of the study.

A distinction may be made between the pilot and feasibility stages of a study (MRC 2000; van Teijlingen and Hundley 2005), although there may be overlap between them and the terms are often used interchangeably. Pilot studies are principally used to test specific aspects of an intervention and the research tools, for example, for validity and reliability (van Teijlingen and Hundley 2005), while feasibility or exploratory studies are used to test the means of delivering the intervention (the study process), at a point when it may be adapted or changed, prior to its implementation in a main trial. Commonly, they are used to provide data from a relevant population to inform the sample size calculation for a subsequent trial, to test trial methods (which may include data collection tools), or to identify the acceptability of a trial intervention within a particular population.

The MRC framework (MRC 2000) describes the exploratory or feasibility study as crucial prior to implementation of the main trial, while Walker et al. (2000)

suggests that key to successful recruitment in a CRT is understanding the structure and organisation of service delivery of the healthcare system in which the trial will operate. In particular, it is important to identify gatekeepers and understand key service changes which may impact on the willingness or ability of the service to participate in a trial. This understanding may be gained during the feasibility stage of a study.

In the current study the preliminary calculations of the sample size suggested that 12 maternity units would be required for the CRT; however, additional data collection was necessary in order to provide a more informed sample size calculation. Prior to conducting the main trial it was also necessary to assess whether a CRT would be feasible within the resources available. For pragmatic reasons such as funding and trial logistics, it was desirable to conduct the study within Scotland. However, rationalisation of maternity services meant that there were likely to be a reduced number of maternity units who would be eligible for trial participation. The feasibility study was required to gauge whether a sufficient number of maternity units in Scotland would be available and willing to participate in a CRT in order to make it possible. Additionally, a feasibility study was required to test the planned trial implementation strategy and to pilot study methods and materials.

7.2. Aim

The aim of study three was to assess the feasibility of conducting a CRT of the use of the algorithm for the diagnosis of active labour in healthy primiparous women, in Scotland.

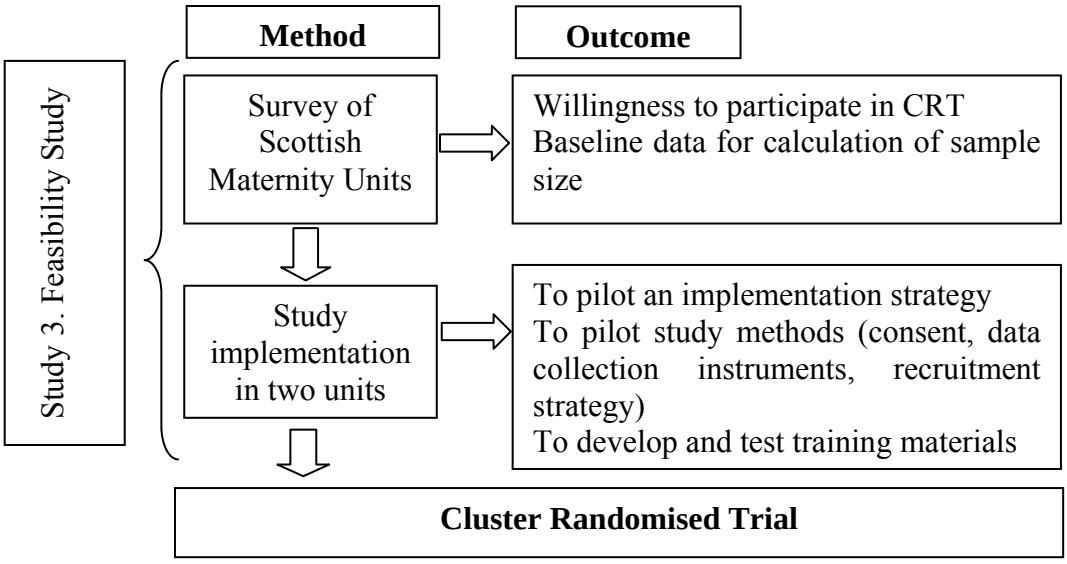
7.3. Objectives:

- To identify maternity units willing to participate in a CRT
- To collect maternity unit level data to inform the CRT sample size calculation
- To pilot the CRT implementation strategy
- To pilot the CRT methods, in particular, the process of gaining consent, the data collection instruments, identification of eligible women and recruitment rates
- To pilot the implementation of the algorithm with particular respect to its acceptability to midwives and identification of training needs.

7.4. Design

This was a feasibility study which used questionnaires, interviews and the CRT methods described in chapter six. Figure 8 provides an overview of methods used in study three. Documentation relating to study three is presented in Appendix 4.

Figure 8 Methods used in study three



7.5. Ethics approval

Ethical approval was granted by the Multicentre Research Ethics Committee for Scotland (MREC/02/10/21) (Appendix 4).

7.6. Sample

7.6.1. Objectives 1 and 2

The sample was Heads of Midwifery of all maternity units in Scotland classified as IIb, IIc and III (Scottish Executive 2002), (n=20). Heads of Midwifery were chosen because they were key gatekeepers in relation to the provision of unit level data and potential CRT involvement.

7.6.2. Objectives 3 to 5

Following the advice of the Multicentre Research Ethics Committee for Scotland it was agreed that it was unnecessary to include both experimental and control

groups in the feasibility study design. Therefore, two maternity units were purposively selected to participate as experimental group sites. Unit one was a large maternity unit classed as level III (Scottish Executive, 2002). The unit had both low and high-risk labour areas, which consisted of labour, delivery and postnatal rooms (LDRP) within four ward areas. Unit two was a small maternity unit classed as IIc (Scottish Executive, 2002). This unit had a traditional style of labour suite comprising of an admission area and individual labour rooms.

7.7. Methods

7.7.1. Objectives 1 and 2

Each Head of Midwifery was sent a questionnaire with a prepaid, addressed return envelope, which they were asked to complete on behalf of their maternity unit. The questionnaire asked whether the unit would be willing (in principle) to participate in a future CRT, as well as a series of questions about unit clinical activity. Where the Head of Midwifery had responsibility for more than one eligible unit she was asked to complete a questionnaire for each or to designate an appropriate senior midwife to complete the questionnaire.

7.7.2. Objectives 3 to 5

Following return of the questionnaires, two units were purposively selected for participation in the feasibility study. The CRT methods were then implemented as described in chapter six, with the exception that no allocation to experimental or control group was required and both sites were asked to use the algorithm.

7.8. Study entry and consent

7.8.1. Cluster level

The strategy for recruitment at cluster level was followed as described in the previous chapter.

7.8.2. Experimental level

A workbook for midwives, containing study information, algorithms and vignettes of case histories, was developed. Training of midwives was implemented as described in chapter six and midwives were then asked for consent to study participation. To assess acceptability of this strategy and the use of the algorithm, semi-structured interviews were conducted with a sample of midwives in each of the maternity units towards the end of the data collection period. Questions focused on their experience of study implementation and training and acceptability of the algorithm.

7.8.3. Observational level

Data routinely collected by the Information and Statistics Division (ISD) of the National Health Service in Scotland was used to estimate the numbers of potentially eligible women prior to the start of the feasibility study. The ISD provides data by health board area therefore local variations in maternity activity are included. It was anticipated that in a maternity unit with 1000 annual births there would be approximately 230 to 250 women per year who would be eligible for the CRT. This was based on the assumption that 45% of women (450) would be primiparous, of these, approximately 7% (32) would be

delivered by elective caesarean section, 30% (135) would have their labour induced, and a further 10% (45) would be excluded due to high risk clinical factors in their pregnancy. The feasibility study aimed to test the validity of these assumptions. During the three month period of data collection it was anticipated that 60 women in each unit would be eligible to participate in the study. The total number of admissions and the number of potentially eligible women not approached for consent were recorded to provide baseline data, an assessment of study compliance and confirmation of the accuracy of the estimated level of recruitment.

Primiparous women in each unit were given information about the study at their antenatal clinic visit between 34 and 36 weeks gestation. On admission to the labour suite eligibility was checked, eligible women received a full explanation of the study, including written information, and were asked to give consent for collection of trial data. Midwives then used the algorithm during their assessment of women on admission, to assist them to diagnose active labour, recording their judgement on the algorithm. There was no control group in the feasibility study.

7.9. Feasibility study outcomes

Outcomes of interest during the feasibility study were:

- Number of units willing to participate in a CRT

- Unit level data: annual births; rates of caesarean section, epidural and oxytocin for augmentation of labour, between unit intrapartum transfers.
- Consent rate of participants at each of the three levels
- Identification of problems with the study implementation strategy and documentation
- Acceptability of the algorithm to midwives

7.10. Data collection analysis

Unit questionnaires were entered onto an Access database then transferred to an SPSS database for analysis. Interviews with midwives were audio recorded and analysed using simple manifest content analysis (Morse and Field 1996). A process by which tapes were listened to repeatedly and responses to specific questions were noted. CRT data including algorithms and trial outcome data (described in chapter six) were collected for all consenting women who fulfilled the study entry criteria, during the three month study period. Data collected were not analysed for CRT outcomes, but were examined for completeness and ambiguous terms, allowing amendment of data collection forms, where necessary.

7.11. Results

7.11.1. Objectives 1 and 2

Questionnaires were returned for all 20 maternity units. All of the units expressed an interest in receiving further information about the study and 90%

(n=18) were willing to be contacted with further information regarding study participation. The remaining two units reported that they were interested in receiving further information but as they were undergoing a process of reassignment to Community Maternity Units they would be unable to participate in the feasibility study or CRT. The characteristics of the units are described on Table 9. Seventeen units had traditional labour suites, one had LDRP rooms (Labour, Delivery, Recovery and Postnatal) only, two units were changing to Community Maternity Units and five units had a midwife managed birth unit in addition to the labour ward.

Table 9 Description of maternity units

Maternity Unit Characteristic	**N =20**	**%**
Type of hospital / unit		
University Teaching	5	25
District General	15	75
EGAMS Level		
IIb	3	15
IIc	13	65
III	4	20
Delivery Suite (not mutually exclusive)		
Labour ward	17	85
Midwife managed birth unit	7	35
LDRP rooms*	4	20
24 hour epidural service	16	80

Unit activity is shown on Table 10.

Table 10 Maternity unit activity

Maternity Unit Activity	N=20	
Annual Births median, range	2222	827 – 5711
Caesarean section rate (overall)	N=20	%
< 20%	4	20
20 - 25%	12	60
26 - 30%	4	20
Emergency caesarean rate (missing data =1)	N=19	
< 10%	2	10
10 - 15%	11	55
16 - 20%	6	30
Epidural rate (missing data =3)	N=17	
<10%	1	5
11-20%	5	25
21-30%	8	40
>30%	3	15
Use of oxytocin in spontaneous labour (missing data = 9)	N=11	
0-10%	4	20
11-20%	6	30
>20	1	5
Transfer of women during labour		
Number of women transferred from index maternity unit during labour (missing data = 3)	N=17	
0	8	40
1-20	6	30
21-40	1	5
41-60	2	10
Number of women transferred to index maternity unit during labour	N=20	
0	11	55
1-20	3	15
21-40	4	20
41-60	0	
>61	2	10

* relates to all unit births

The information available varied. Although all units were able to identify their caesarean section rate, only 11 were able to provide information on the use of oxytocin in spontaneous labour and these data included both primiparous and multiparous women. The emergency caesarean section data confirmed the preliminary calculation (using the ISD figures) for the ICC for emergency CS of

0.005. However, the data provided by the questionnaires provided insufficient information to revise the ICC for oxytocin use in primiparous women.

The 18 units who indicated their willingness to participate in the CRT were contacted by telephone and asked to provide further data on the use of oxytocin for augmentation of labour specifically in primiparous women. In some units these data could be extracted from routinely collected, computerised data, while in others data required to be collected retrospectively, by hand searching, for a sample period. Twelve units were able to provide some data. These data did allow a calculation of the ICC for use of oxytocin and the recalculation of the sample size for the CRT as follows.

7.11.2. CRT sample size calculation

Data were collected on the number of births to primiparous women and the number of these women who received oxytocin for augmentation of spontaneous labour for 12 maternity units in Scotland who were able to provide data. The mean period of data collection was 3.8 months (range 1 to 12). In this sample, the mean proportion of women given oxytocin for augmentation of labour was 34%, mean deliveries per month was 61 (SD = 40), the ICC was 0.041, and the SD between hospitals was 0.096. A difference of 10% in the proportion of women who received oxytocin for augmentation of labour was deemed clinically relevant. To detect a difference between 34% and 24% with power 0.90, 431 women per group would be required in an individually

randomised study. This required to be adjusted for the clustering by maternity unit by considering the design effect as follows.

Assuming that 12 maternity units would be available with a total of 732 births per month, and the same number of births in each unit (cluster); the following calculation was used to determine the number of births required to achieve the power which would be obtained using an individual randomised design. First, assuming that units were weighted by the number of deliveries, the design effect was calculated by $1+(m-1)\rho$, where m is the number of births in each cluster and ρ is the ICC. The effective sample size was then obtained by dividing the sample size by the design effect. This is the sample size in an individually randomised study of the same power as the CRT (ICC = 0.041). For m = 200 deliveries per hospital, the design effect would have been 9.16 and the effective sample size 131 women per group. Increasing the number of deliveries per hospital increased both the number of women and the design effect, therefore the effective sample size increased very slowly as shown in Table 10.

Table 11 The design effect

Births per unit	Design effect	Number of births per group	Effective sample size
50	3.01	300	99.7
100	5.06	600	118.6
150	7.11	900	126.6
200	9.16	1200	131.0
300	13.26	1800	135.8
400	17.36	2400	138.3
500	21.46	3000	139.8
600	25.56	3600	140.9

This demonstrated that the required effective sample size could not be reached. The number of potentially available maternity units could not be substantially increased, if the trial was to be conducted within Scotland as planned. Therefore, the effect of variation between hospitals required to be reduced. This was achieved by incorporating baseline data for the cluster; collecting data for the same number of women before and after study implementation and using the proportion of women receiving oxytocin before trial implementation as a covariate in a hospital level analysis.

A regression (or covariance) analysis on baseline was planned. In order to obtain the appropriate sample size for a hospital level analysis of covariance, the correlation between proportions of women given oxytocin before and after the intervention was required. Using data for 200 women before and 200 women after the time of the intervention in each hospital (including control hospitals), it was estimated that the correlation would be 0.89 and the standard deviation of the proportions would be 0.10. This was done by the study statistician using simulation. These estimates were used in the Stata 8 sampsi command to estimate a study power of 0.97 for detecting the difference between 34% and 24% oxytocin use, at significance level 0.05 'after' oxytocin use, using a total of 12 hospitals. Therefore, the aim was to recruit a minimum of 12 maternity units (the statistical power of the study would increase if more units were included). Within each unit the target cluster size was 400 women, comprising 200 before and 200 after the point of study implementation. Anonymous data would be used for the baseline sample.

7.11.3. CRT analysis

Changing the sampling strategy to include baseline data collection had implications for the analysis of the trial data as follows. Analysis of data was appropriate for CRTs and clustering of observations within maternity units was accounted for. All analyses were done using Stata 8 (Stata Corp., College Station, Texas). Cluster level analysis controlled for baseline was used. This meant that for each outcome a summary statistic (the mean or proportion) was calculated for each cluster, at baseline and after study implementation. In each case the baseline value was the covariate. For example, for the primary outcome (use of oxytocin), the proportions of women receiving oxytocin at baseline and following study implementation was calculated for each unit. Regression analysis was then conducted. Regression allows the prediction of one variable from another. In this case the proportion of oxytocin use after study implementation was predicted by the baseline proportion. Therefore regression was conducted of the final proportion on the baseline proportion and the treatment group (experimental and control). This analysis takes into account variation in cluster size, and provides an estimated difference in percentage for use of oxytocin (intervention group minus control group), the confidence interval (CI) and test of significance for the difference in proportions of women receiving oxytocin. Analysis for secondary outcomes was conducted in the same way. In order to ensure that analysis was hypothesis driven a data analysis plan was developed before analysis was conducted (Appendix 5).

7.11.4. Objective 3. Implementation Strategy

As the initial point of contact for each unit the Heads of Midwifery in units one and two received a presentation and written information about the study. Both agreed to participate and each chose to discuss the study with, and seek consent for unit participation from, the Clinical Director rather than have a direct approach by the research team.

Unit one served an urban area and had 4675 annual births. The unit had four wards each with LDRP rooms. One hundred and twenty midwives worked in these wards and could be involved in providing intrapartum care. All required to be approached for study participation. A presentation was made to midwifery team leaders and this was followed by regular short workshops (24 in total) and individual contacts with midwives at ward level. Unit two had 1090 annual births, serving a more rural area. The unit had one labour ward with a core group of 38 midwives who provided intrapartum care. A meeting was held with senior midwives to discuss the study and this was followed by six workshops for midwives.

7.11.5. Objective 4. Pilot of data collection instruments, identification of eligible women, recruitment and consent rates

Study compliance varied between the units, with limited baseline data being available for unit one. There were 1057 admissions to unit one during the study period, of which approximately 43% (455) were primigravid women at term. However, no information was available on the number of women who were

eligible for study entry but were not approached. Forty-one women were approached for study participation of these 36 consented (89% consent rate).

In unit two there were 248 admissions during the study period. Of these 113 (45%) were primigravid women at term, 40 women were ineligible for study entry for clinical reasons (e.g. induction of labour), leaving 73 women who were approached for study entry and of whom 60 consented (82%). Identification of eligible women and the high consent rate in this unit confirmed initial estimates for potential participation which suggested that for a unit of approximately 1000 births per year a sample size of 200 could be obtained in a 10-month period. Consent rates for each level of participation in the feasibility study are shown in Table 12.

Table 12 Consent rates for feasibility study participation

Level of consent				Total approached	Consent to participation	Consent rate
Cluster; Maternity Unit				2	2	100%
Experimental; Midwives						
Unit one (n=120)				120	67	56%
Unit two (n=38)				38	29	76%
Observation; Women						
	All births	Prim (%)	eligible			
Unit one	1057	454 (43%)	N/A	41 (% unknown)	36	89%
Unit two	248	113 (46%)	73	73 (100%)	60	82%

N/A = not available

In both units there were regular visits and telephone contacts by the research midwife to the local study co-ordinator throughout the data collection period. These focussed on issues of study compliance and data collection. Strategies for improvement in compliance were discussed including information letters and posters for staff areas, provision of study pens and biscuits for staff.

7.11.6. Objective 5. Acceptability of the algorithm and identification of training needs

Interviews were carried out with six midwives (three from each unit), the interviews were conducted by the author of this thesis. Two of the midwives interviewed had attended a workshop presentation; both felt that this was useful in providing information about the study. Three of the midwives had first heard about the study from a colleague who had attended a workshop and were then given a study workbook; all felt that this was an appropriate means of receiving information about the study. One midwife first heard about the study through reading information leaflets for women, she felt that this was inadequate, and would have preferred to attend a workshop. Three of the midwives felt that more workshop sessions would have been beneficial. All the midwives reported that the information contained in the workbook provided good information about the study. None of the midwives had experienced any difficulty in completing the study paperwork or the algorithm, which they reported was straightforward and quick to complete.

Interviews with midwives from unit one highlighted a number of issues relating to the implementation of the study in their unit. Because of the nature of the organisation of maternity care in that unit, women in labour could be admitted to any one of four ward areas, each of which also provided antenatal and postnatal care. This meant that there was no central point for co-ordinating the study at the point of admission for women in labour. This differed from unit two where there was a central point for study entry as all women were admitted to one labour ward.

Midwives also highlighted issues surrounding the nominated study co-ordinator in unit one. Firstly, a midwife on a management secondment had been nominated, she was not clinically based and this was negatively perceived. Subsequently, at the mid-point of data collection, two clinical midwives were nominated however, they were perceived to be too junior to effectively promote the study.

All of the midwives interviewed reported that they were happy to participate in the study and felt that it was generally acceptable to midwives in the units. One of the midwives reported that she felt that the algorithm was an excellent tool in particular, for teaching inexperienced midwives. Another reported that she felt that the algorithm was an excellent idea, which would help midwives to focus care on women who were actually in labour rather than on those who would be better at home.

7.12. Discussion

7.12.1. Availability of units

All heads of midwifery surveyed expressed an interest in taking part in a CRT this was a very positive response and suggested that a CRT, conducted in Scotland, could be feasible. However, an ongoing process of reorganisation of maternity services in Scotland meant that during the course of the feasibility study, two potentially eligible maternity units were reclassified as Community Maternity Units, thus reducing the pool of available units for the CRT.

7.12.2. Data collection

Following recalculation of the required sample size data was required on 200 women at baseline and after study implementation in each unit. The data collection strategy required to be changed accordingly. A study start date was first agreed with each maternity unit (experimental and control). Local trial co-ordinators were asked to identify, retrospectively, a sample of 200 women who had given birth prior to the start date and who would have been eligible for the study prior to the onset of labour. Anonymous trial outcome data were collected for these women, this formed the baseline sample. Recruitment and collection of data for women after the study start date remained as described in the previous chapter.

7.12.3. Pilot of RCT implementation strategy and methods

There was considerable variation in study compliance across the two units. One of the issues highlighted was the organisational structure of unit one,

which meant that there was no central point for the co-ordination of the study. Unit one was the only unit in Scotland with an organisation system that had only LDRP rooms. All other units had designated labour wards, which would provide a central point for study co-ordination. The second issue was the choice of midwives to co-ordinate the study at local level, midwives were perceived as being either not clinically focused or too junior. This highlighted the importance of identifying a midwife who was currently practicing in the delivery area and sufficiently senior to have the respect of other midwives. Accordingly the aim was for all local trial co-ordinators for the CRT to be senior clinical midwives.

There was a lack of accurate information on the number of potentially eligible women who were not approached for consent to study entry. This was a particular problem in unit one. Although a data collection form was used to collect the total number of admissions and the number of potentially eligible women not approached for consent, this was not found to be effective in the larger maternity unit. Additionally, it became clear that an accurate record of eligible women could not be made without a review of the case records of all primiparous women admitted in spontaneous labour, which would identify individual exclusion criteria (for example hypertension). This would have been very resource intensive and prohibitive in terms of time required for data collection, therefore routinely collected central data sources (ISD) were used to provide estimates of potentially eligible women in the CRT.

7.12.4. Consent

The training package for midwives (workshops and workbooks) were developed and tested. The interviews indicated that both strategies were important in ensuring that midwives understood the nature of the study. Although the midwives interviewed reported that the algorithm was an acceptable and useful tool the consent rate was low in unit one where only 54% of midwives consented. There was also low compliance with study protocol in this unit resulting in recruitment of only 36 women. This would suggest that many of the midwives were reluctant to take part in the study and that the midwives interviewed were merely giving socially acceptable answers. However, midwives who had not themselves signed consent forms did, in some cases, recruit women to the study and complete algorithms (this could be seen from midwives counter signatures on women's consent forms). Feedback from midwives suggested that they frequently took part in research (and collected data for clinical audit) without being asked specifically to consent. It may have been that the midwives did not realise that they were participants in the study rather than merely collecting study data. Increased emphasis was placed on the importance of signing a consent form for study participation in the CRT. The high consent rate obtained from women who were approached, in both units, indicated that the method of consent gaining and the study itself was acceptable to women.

7.13. Summary

A number of authors have highlighted the value of conducting a feasibility study prior to a full trial for example MRC (2000), van Teijlingen and Hundley (2005)

and Walker et al. (2000). The information and experience gained in conducting this feasibility study supports this. The study provided information for the CRT sample size calculation and importantly it highlighted the need for a change in recruitment strategy to provide baseline data. It identified the number of maternity units which would be willing (in principle) to participate in a CRT, and allowed the identification of key gatekeepers (for example, Heads of Midwifery, Clinical Directors, labour suite managers, records officers) within each maternity unit. This information was essential to the smooth running of the subsequent CRT.

The CRT implementation strategy was tested and found to be successful, in particular at the level of the cluster and the observational level. The strategy at the experimental level was equivocal with more success achieved in one unit than the other however, adjustments were made to the strategy which aimed to improve recruitment and study compliance at this level. Workbooks for midwives and all other study materials were developed and successfully tested.

There were limitations to the feasibility study. A control group was not included in this stage therefore the data collection strategy was not tested for this group. Further this stage of the study did not test the algorithm in respect of its efficacy in improving clinical outcomes for women. The following chapter reports the results of the CRT which aimed to test the efficacy of the algorithm in improving clinical outcomes for women.

CHAPTER 8: STUDY FOUR: CLUSTER RANDOMISED TRIAL – RESULTS

In this chapter the results of the CRT are presented following the framework suggested for the reporting of CRTs (Elbourne and Campbell 2001: Campbell et al. 2004). This framework provides guidance for the extension of the CONSORT statement, originally developed to improve the reporting of randomised controlled trials, to CRT designs. Documents relating to study four are presented in Appendix 5.

8.1. Participants

8.1.1. Cluster level

The trial was conducted between March 2005 and June 2007. During this period 15 maternity units in Scotland were eligible to participate. Of these, 14 consented and were allocated to experimental (n=7) or control (n=7) groups. One unit declined to participate because of other planned research commitments. Once entered, all units completed the trial as allocated. Baseline descriptive data for each cluster is presented in Table 13. Data are presented for experimental and control groups for the following characteristics; number of annual births, presence or absence of a MBU, percentage of births to primiparous women living in the most deprived areas (as defined by the Scottish Index of Multiple Deprivation) and unit type (as defined by Scottish Executive, 2002). Most of the units in both groups were classified as 11c (Scottish Executive, 2002), and annual births ranged from 950 to 5242. Two units in each group had an onsite MBU.

Table 13 Baseline characteristics of clusters

	Unit No	Total Annual Births	MBU	% in most deprived areas	Unit type (EGAMS)
Intervention	2	3166	No	14.3	11c
	4	1305	No	12.2	11c
	7	3324	Yes	15.4	11c
	9	1888	No	7.1	11c
	10	950	No	0.5	11b
	12	5242	Yes	47.1	111
	14	3535	No	20.9	111
Control	1	1042	No	2.5	11b
	3	2988	No	31.9	11c
	5	4183	Yes	6.9	111
	6	3426	No	36.5	111
	8	3590	No	28.7	11c
	11	2710	Yes	31.4	11c
	13	2743	No	12.8	11c

8.1.2. Experimental level

Overall, 80% of midwives consented to participate (unit range 57-100%); one unit lost all completed midwife consent forms and is therefore excluded from the consent rate.

Table 14 shows the labour suite complement of midwives and the number of midwives who consented to trial participation by cluster. Recruitment appears to exceed the total number of midwives in one cluster because a team model of midwifery care was operating in that unit. This meant that most hospital and community based midwives had a labour suite commitment during the study period.

Table 14 Midwife number and consent by cluster

Unit No	Delivery suite midwives	Consent No.	% Consent
2	30	102	100
4	48	31	65
7	27	Missing data	Missing data
9	33	31	94
10	39	25	64
12	61	35	57
14	26	24	92

8.1.3. Observational level

The flow of participants by cluster is shown on Table 15. This identifies the number of potentially eligible women, data collected at baseline and after study implementation. The number of potentially eligible women was taken from ISD data and estimates the number of women who would have been potentially eligible for recruitment during the planned ten month data collection period after trial implementation in each cluster. The smallest units (in annual births) did not necessarily have the fewest number of women who were potentially eligible for the trial. Units were requested to recruit 200 women after trial implementation, but only nine units managed this. Although the data on potentially eligible women are estimates, they suggest that the smallest units may have experienced difficulty in recruiting the target sample within the trial period.

Table 15 Observation level data at baseline and after study implementation

	Unit No	Potentially eligible women*	Baseline	After study implementation (target 200)	Data Analysed
Experimental	2	642	198	200	200
	4	138	48	65	64
	7	731	83	57	56
	9	355	202	200	199
	10	156	200	60	60
	12	578	162	200	200
	14	550	136	114	113
Total		3150	1029	896	892
Control	1	248	201	200	200
	3	402	199	200	199
	5	842	199	200	200
	6	348	200	200	200
	8	769	199	200	200
	11	538	197	200	200
	13	555	96	81	80
Total		3702	1291	1287	1279

*ISD 2006 & SPCERH 2003

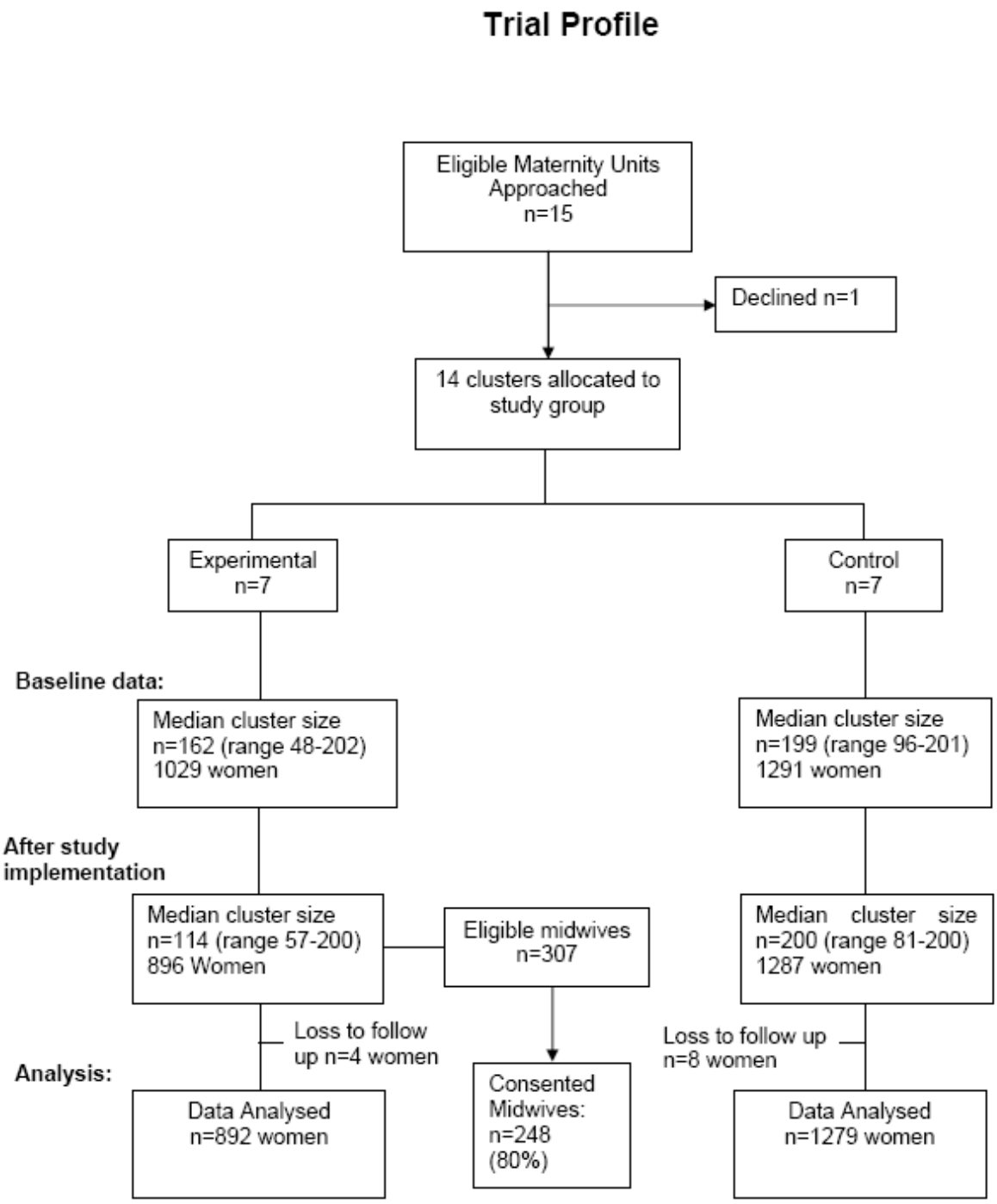
Retrieval of archived case records, for baseline data collection, was more difficult than anticipated in some units (despite financial provision for payment to hospital records departments). Only four units achieved the requested 200, although five others came very close. Barriers included maternity unit policy regarding retrieval of records for research and storage of archived records off site. This resulted in a few units being unable to complete data collection at baseline. Ultimately baseline data were collected for 1029 women in the experimental group and 1291 women in the control group.

Monitoring of early returns of data revealed that recruitment was slower than expected in control units and because women were recruited in the early postnatal period, the possibility of recruitment bias was recognised (e.g.

midwives may have only approached women with good clinical outcomes for study consent), this problem was not anticipated as control sites were not included in the feasibility study. The data collection strategy was therefore changed as follows. Midwives in control units collected baseline data as planned; they then sought consent for data collection from 100 women after trial implementation (this was the number required for the postnatal questionnaire for the health economics evaluation) and 100 contemporaneous, anonymised cases (making a total of 200 cases). This resulted in near complete data collection in control units. However, this strategy could not be used in experimental units as informed consent was required from all women on admission to labour suites. Data collection was extended in five units to increase recruitment. There was a small amount of missing data in the 'after study implementation' period (Experimental group n= 4; Control group n= 8), which was due to the inability to retrieve case records in a few cases. Following trial implementation complete data were collected for 896 women in the experimental group and 1279 women in the control group. All women recruited were eligible for the trial (Appendix 5) at first labour suite assessment.

The trial profile is shown in Figure 9, this summarises the flow of participants through each stage (cluster, experimental and observational), showing the median cluster size and range at baseline, after study implementation and at data analysis. There was variation in cluster size at baseline and after study implementation, in particular in the experimental group. This was accounted for in the subsequent regression analysis.

Figure 9 Trial Profile



8.2. Primary outcome

8.2.1. Oxytocin for augmentation of labour

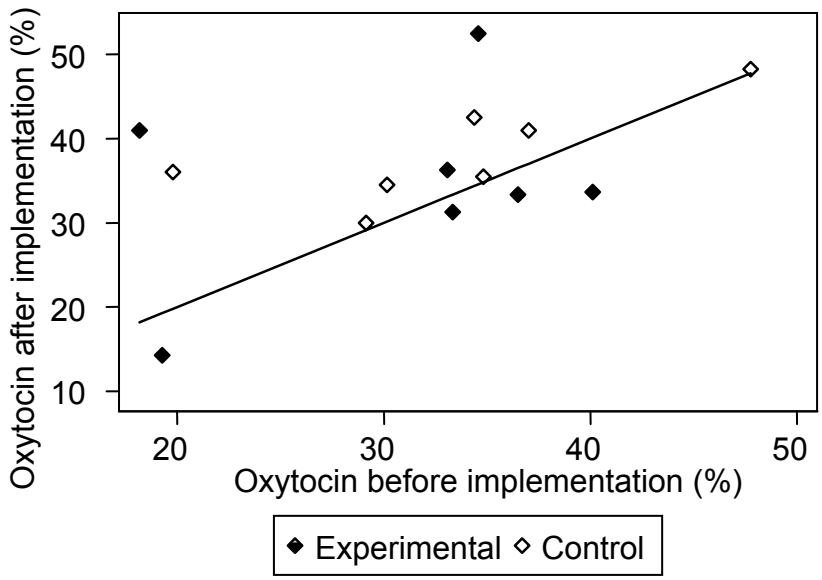
The proportion of women given oxytocin for augmentation of labour was calculated for each cluster at baseline and following trial implementation (Table 16).

Table 16 Oxytocin use at baseline and after study implementation

Unit No	Total women per cluster (before & after)	Oxytocin use at baseline (n)	% Oxytocin use after study implementation (n)
Experimental	n=1921		
2	398	18.2 (36)	41.0 (82)
4	112	33.3 (16)	31.3 (20)
7	139	19.3 (16)	14.3 (8)
9	401	40.1 (81)	33.7 (67)
10	260	36.5 (73)	33.3 (20)
12	362	34.6 (56)	52.5 (105)
14	249	33.1 (45)	36.3 (41)
Control	n= 2570		
1	401	34.8 (70)	35.5 (71)
3	398	47.7 (95)	48.2 (96)
5	399	29.1 (58)	30.0 (60)
6	400	37.0 (74)	41.0 (82)
8	399	30.2 (60)	34.5 (69)
11	397	19.8 (39)	36.0 (72)
13	176	34.4 (33)	42.5 (34)

These data are shown graphically in Figure 10.

Figure 10 Proportion of women receiving oxytocin before and after trial implementation



Points above the line show clusters where the proportion of women given oxytocin increased after trial implementation, points below the line show clusters where the proportion decreased post implementation. Control sites are consistently above the line, while intervention sites are on both sides of the line.

Regression analysis (Figure 11) was conducted of the percentage oxytocin use after trial implementation on the baseline percentage, and by study group (experimental minus control).

Figure 11 Regression analysis for use of oxytocin

```
regress  oxyaft oxybef inter [aweight=n]
(sum of wgt is   4.4910e+03)

      Source |       SS       df       MS              Number of obs =      14
-------------+------------------------------           F(  2,    11) =    0.96
       Model |  120.150321     2  60.0751607           Prob > F      =  0.4137
    Residual |  690.145046    11  62.7404588           R-squared     =  0.1483
-------------+------------------------------           Adj R-squared = -0.0066
       Total |  810.295368    13  62.3304129           Root MSE      =  7.9209

------------------------------------------------------------------------------
      oxyaft |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      oxybef |   .3549639   .2570645     1.38   0.195    -.2108312     .920759
       inter |   .3140866   4.309124     0.07   0.943    -9.170232    9.798405
       _cons |   26.09049   8.982563     2.90   0.014     6.320003    45.86098
------------------------------------------------------------------------------
```

oxybef = oxytocin use before trial
inter = intervention
cons = constant

This gave the estimated difference in percentage of women given oxytocin (adjusted for baseline) of 0.3, p = 0.9, 95% CI –9.2 to 9.8. This indicates that there was no significant difference in percentage of women who received oxytocin attributable to the use of the algorithm.

8.3. Secondary outcomes

8.3.1. Intervention in labour

Four labour interventions were considered; these were artificial rupture of membranes (ARM), continuous electronic fetal monitoring (EFM), use of pain relief and vaginal examination (VE) (Table 17). For table 17 and subsequent tables, summary descriptive data are presented for experimental and control groups at baseline and after trial implementation, these data do not take account of the effects of clustering. Data for each outcome were analysed using regression as described above (this analysis takes account of clustering).

The percentage difference (intervention minus control) has been adjusted for baseline value. There was no significant difference between groups for any of the labour interventions.

Table 17 Interventions in labour

Item	Experimental n (%)		Control n (%)		% Diff	p value	95% CI
	before n=1029	after n=892	before n=1291	after n=1279			
Intervention							
ARM	383 (37.2)	401 (44.9)	514 (39.8)	500 (39.0)	5.6	0.1	-2.2 to 13.4
Continuous EFM	567 (55.1)	557 (62.4)	781 (60.4)	820 (64.1)	-0.1	1.00	-14.2 to 14.1
Epidural	211 (20.5)	290 (32.5)	382 (29.5)	441 (34.4)	2.1	0.7	-8.0 to 12.2
Opiate	646 (62.7	532 (59.6)	680 (52.6)	649 (50.7)	1.5	0.6	-4.6 to 7.6
Epidural & opiate	129 (12.5)	177 (19.8)	223 (17.2)	225 (17.5)	4.4	0.2	-2.8 to 11.7
VE mean (range)	2.89 (0-11)	3.67 (0-11)	3.31 (0-10)	3.46 (0-11)	0.2	0.3	-0.3 to 0.7

8.3.2. Admission management

Outcomes relating to admission management were; the number of labour suite admissions, the length of time from admission to delivery and the length of time in active labour (Table 18).

Table 18 Number of admissions

Item	Experimental		Control		% Diff	p value	95% CI
	before n=1029	after n=892	before n=1291	after n=1279			
Admissions n (%)							
One admission	617 (60.0)	398 (44.6)	798 (61.8)	795 (62.6)	-19.2	0.002	-29.9 to -8.6
Admissions prior to labour							
Mean admissions (range)	1.28 (1-4)	1.45 (1-4)	1.26 (1-4)	1.28 (1-6)	0.29	0.03	0.04 to 0.55
Number							
1	308 (29.9)	305 (34.2)	382 (29.6)	366 (28.6)			
2	79 (7.7)	149 (16.7)	85 (6.6)	88 (6.9)			
3	14 (1.4)	32 (3.6)	16 (1.2)	17 (1.3)			
4+	2 (0.2)	3 (0.3)	3 (0.2)	3 (0.2)			
Missing data	9	5	7	10			

Significantly more women in the control group had only one admission, that is, they were more likely to remain in the labour suite until delivery following their first admission assessment (% diff = -19.2, p= 0.002, 95% CI -29.9 to -8.6), while women in the experimental group had significantly more admissions resulting in discharge from the delivery suite prior to delivery (% diff = 0.29, p= 0.03, 95% CI 0.04 to 0.55).

The time spent in labour is presented in two ways; the mean time (in hours) from final admission to delivery and mean duration of active labour (Table 19). These data were subject to a number of errors and missing information, these were corrected where possible and where not possible were recoded as missing data. There was no difference between groups either for duration of active labour, or time from admission to delivery. Mean duration of active labour exceeded the mean time from admission to delivery because active labour commenced prior to admission in a number of cases.

Table 19 Time in labour ward and duration of active labour

Item	Experimental mean (SD)		Control mean (SD)		% Diff	p value	95% CI
	before n=1029	after n=892	before N=1291	after n=1279			
Time in labour							
Admission to delivery	8.08 (5.68)	9.60 (11.29)	7.81 (5.07)	8.06 (5.41)	0.75	0.2	-0.55 to 2.05
Missing data (n)	109	51	98	77			
Duration of active labour	9.91 (5.35)	10.82 (5.52)	9.55 (4.96)	9.54 (5.17)	0.61	0.2	-0.45 to 1.67
Missing data (n)	145	69	112	98			

Table 20 shows the mean time from final admission to delivery and duration of active labour, by the number of admissions prior to labour. Only descriptive data are provided because the number of prior admissions is an outcome variable. It would not have been appropriate to stratify data by an outcome variable in order to compare groups.

Table 20 Length of labour by number of admissions prior to labour

Item		Experimental n (%)		Control n (%)	
		Before n= 1029	after n= 892	before n= 1291	after n= 1279
Time in labour mean (SD)					
Admission to delivery					
Prior admissions	0	7.72 (5.09)	9.61 (6.02)	7.30 (5.01)	7.81 (5.43)
	1	8.50 (5.81)	8.50 (5.42)	8.57 (5.07)	8.38 (5.30)
	2	8.46 (4.83)	11.52 (23.77)	9.92 (4.84)	8.55 (5.36)
	3	13.00 (14.55)	9.57 (3.95)	8.11 (5.07)	10.46 (6.61)
	4	0	11.82 (11.65)	4.45 (2.99)	12.24 (3.83)
	5	0	0	0	0
	6	0	0	0	12.90
Missing data		109	51	98	77
Duration of active labour					
Prior admissions	0	9.74 (5.10)	11.31 (5.67)	9.19 (4.88)	9.50 (5.14)
	1	10.37 (5.77)	10.47 (5.69)	9.93 (5.06)	9.31 (4.97)
	2	9.67 (5.69)	10.10 (4.73)	11.46 (4.75)	10.32 (6.12)
	3	9.34 (6.20)	11.67 (5.25)	10.79 (5.45)	11.06 (5.00)
	4	0	6.15 (2.7)	10.76 (4.48)	14.61 (2.23)
	5	0	0	0	0
	6	0	0	0	12.90
Missing data		145	69	112	98

8.3.3. Mode of delivery

There was no difference in mode of delivery between study groups (Table 21).

Table 21 Mode of delivery

Item	Experimental n (%)		Control n (%)		% Diff	p value	95% CI
	before n=1029	after n=892	before n=1291	after n=1279			
Mode of delivery							
SVD	709 (68.9)	526 (58.9)	810 (63)	785 (61.3)	-3.2	0.6	-15.1 to 8.7
Breech	3 (0.2)	0	0	0			
Instrumental	205 (19.9)	241 (27.0)	319 (25)	323 (25.2)			
Elective C/S	4 (0.3)	0	0	3			
Emergency C/S	106 (10.3)	123 (13.7)	162 (12.5)	165 (12.9)	0.0	1.0	-4.3 to 4.3
Missing data	2	2	0	3			

The trial entry criteria excluded breech presentation and elective caesarean births, however, in three cases breech presentation was not diagnosed on admission. In a further seven cases the decision was made to perform a semi-elective caesarean section following the initial labour suite assessment (and trial entry) resulting in subsequent discharge home. As these women were eligible at the point of trial entry they were not excluded from data analysis.

8.3.4. Neonatal outcomes

The following neonatal outcomes were considered (Table 22); Apgar score (mean score at one and five minutes and score of less than seven at five minutes), resuscitation (excluding mucus extraction only), admission to the neonatal unit (NNU) and birth before arrival at the planned maternity unit (BBA).

Table 22 Neonatal outcomes

Item	Experimental		Control		% Diff	p value	95% CI
	before n=1029	after N=892	before n=1291	after n=1279			
Neonatal outcome Mean (SD)							
Apgar at 1 min	9.52 (10.2)	8.84 (6.9)	8.97 (7.24)	9.21 (8.85)	-0.00	0.9	-0.17 to 0.15
Apgar at 5 min	9.25 (0.70)	9.27 (0.79)	9.10 (0.74)	9.14 (0.69)	-0.08	0.4	-0.27 to 0.11
Apgar< 7 at 5 min n (%)	7 (0.6)	9 (1.0)	18 (1.3)	13 (1.0)			
Resuscitation	130 (14.0)	106 (12.7)	151 (12.0)	145 (11.6)	-0.9	0.7	-6.4 to 4.7
Missing data	101	58	29	29			
Admitted to NNU	38 (3.6)	29 (3.2)	56 (4.3)	60 (4.6)	-0.4	0.7	-2.6 to 1.8
BBA	6 (0.5)	11 (1.2)	9 (0.6)	11 (0.8)			

Overall 67 babies were admitted to the neonatal unit for more than 48 hours. There were very few unplanned out of hospital births (BBA) or babies with an Apgar score less than seven at five minutes; therefore no statistical analysis was conducted for these variables. There was one stillbirth in the control group at baseline.

8.3.5. Maternal complications

Maternal complications are shown in Table 23. Overall 45% of women (n= 2,028) experienced at least one intrapartum complication. Statistical comparison was made only for complications which occurred in at least 100 cases. These were; the occurrence of any complication, failure to progress in the first stage of labour, failure to progress in the second stage of labour, fetal distress (by electronic monitoring) and meconium stained liquor. Descriptive data are presented for complications which occurred in at least 10 cases overall.

Table 23 Maternal complications

Complication	Intervention n (%)		Control n (%)		% diff	p value	95% CI
	before n=1029	After n=892	before n=1291	After n=1279			
Any complication	422 (41)	439 (49.2)	571 (44.2)	596 (46.6)	3.9	0.5	-9.4 to 17.2
Failure to progress 1st stage	70 (6.8)	42 (4.7)	55 (4.3)	59 (4.6)	-3.4	0.5	-15.3 to 8.6
Failure to progress 2nd stage	91(8.8)	142 (15.9)	84 (6.5)	119 (9.3)	15.2	0.1	-4.5 to 34.9
Fetal distress	152 (14.7)	166 (18.6)	245 (19.0)	242 (18.9)	2.4	0.6	-6.6 to 11.3
Meconium stained liquor	152 (14.8)	133 (14.9)	213 (16.5)	211 (16.6)	-0.5	0.9	-7.2 to 6.3
Mal position/ presentation	11 (1.1)	9 (1.0)	10 (0.8)	16 (1.2)			
Intrapartum haemorrhage	10 (1.0)	5 (0.5)	6 (0.5)	7 (0.5)			
Post partum haemorrhage	12 (1.2)	10 (1.1)	16 (1.2)	20 (1.5)			
Failed forceps	4 (0.4)	9 (1.0)	1 (0.1)	3 (0.2)			
Shoulder dystocia	4 (0.4)	5 (0.5)	13 (1.0)	7 (0.5)			
Maternal pyrexia	2 (0.2)	3 (0.3)	12 (0.9)	10 (0.7)			
Raised blood pressure	5 (0.5)	4 (0.4)	5 (0.4)	6 (0.4)			
Retained placenta	11 (1.1)	16 (1.7)	26 (2.0)	14 (1.0)			
3rd/4th degree tear	8 (0.8)	7 (0.7)	10 (0.8)	8 (0.6)			

There was no significant difference between groups for maternal complications. There were wide between cluster variations for the number of women who were recorded as having failure to progress in the second stage of labour, occurrence ranged from less than 1% to 20%. This was the only complication which showed such wide variation and this may have been due to different definitions used for failure to progress in the second stage between units.

8.4. Summary

The study aimed to recruit 12 units, ultimately recruiting 14. Although some sites were unable to recruit the planned sample of women, this was partially

offset by the two extra units recruited at cluster level. All units completed the study as per unit of allocation.

There was no difference in proportion of women who received oxytocin for augmentation of labour or for other interventions in labour. However, significantly more women in the control group remained in the labour ward following their first admission, while women in the intervention group were more likely to be discharged home. This did not significantly reduce overall time spent in the labour ward or duration of active labour.

CHAPTER 9: DISCUSSION

The CRT which is presented in this thesis was the first adequately powered trial of the use of diagnostic cues for active labour on clinical outcomes for women. The trial involved 14 maternity units and 4503 women and tested the effectiveness of an algorithm to assist midwives with the diagnosis of active labour in primiparous women. The results showed that use of the algorithm did not reduce the number of women who received oxytocin or other medical interventions compared with standard care. Significantly more women in the control group remained in the labour ward until delivery following their first admission, while women in the experimental group were more likely to be discharged home and subsequently had significantly more admissions prior to labour. There was no significant difference between groups for maternal or neonatal complications or unplanned out of hospital births. There was no significant difference between groups in the time spent in the labour suite during labour or in the duration of active labour.

Although the results demonstrated that there was no significant difference in the primary outcome, the strength of the study design means that it contributes significantly to the debate on early labour management. The diagnosis of labour has important clinical and resource implications for the care of a woman in labour, yet only one other trial has tested whether gate keeping the admission of women to labour wards, by improving the accuracy of labour diagnosis would lead to reduced interventions in labour (McNiven et al. 1998). In contrast to the findings of the CRT presented in this thesis, the McNiven et

al. (1998) trial reported that when labour was assessed using strict diagnostic criteria, significantly fewer women received oxytocin to augment labour and less pain relief was used compared to no labour assessment (22.9% compared to 40.4% and 7.6% compared to 20% respectively). Although both trials included similar diagnostic criteria, the interventions were not identical. In the McNiven et al. (1998) study low risk women were randomly allocated to study groups when they presented in spontaneous labour. All women in the control group were then admitted directly to the labour ward without prior labour assessment, while women in the experimental group had their labour assessed using the diagnostic criteria; those judged not to be in labour were sent home or remained in an assessment area to await the establishment of active labour before being admitted to the labour ward. Thus, McNiven et al. (1998) evaluated a package of care which included both diagnosis and management of early labour while in the studies presented in this thesis the judgement and decision-making process was deconstructed in order to study one clearly defined element, the diagnostic judgement. In the CRT presented in this thesis midwives in both experimental and control groups carried out a labour assessment of women prior to the labour ward admission. In the experimental group this assessment was supported by the algorithm which provided a recommendation that women diagnosed as not being in active labour would be discharged home or admitted to an antenatal area. However, in both groups, decisions about clinical management (i.e. whether to admit or discharge a woman) were ultimately determined by the midwives. Thus, the groups in the CRT differed only in the use of the algorithm and were therefore likely to be more similar than the groups in the McNiven et al. (1998) study.

There are advantages and disadvantages to evaluating a package of care, or a single intervention. In this situation evaluating a package of care may be more pragmatic in that it treats labour assessment and admission management as one intervention (a truly complex intervention (MRC 2000)). However, it is then not possible to identify what the active ingredient in the experimental group was. Alternatively, although decision support tools such as the algorithm are classed as complex interventions by the MRC (2000), they are in practice, relatively simple. Implementing a trial of a single intervention may mean that there are a number of unanswered questions and further study may be required. However, the advantage is that the active ingredient in the trial is clearer. Consequently the results of the CRT are more likely to be an accurate estimate of the effect of using explicit diagnostic cues for diagnosis of active labour on the rate of oxytocin use.

A further difference between the McNiven et al. (1998) trial and the CRT reported in this thesis is in the design and scale of the studies. McNiven et al. (1998) conducted a study in one hospital and included only 209 women. The study was therefore underpowered to report a statistically significant difference in a number of important clinical outcomes and may have been affected by contamination between study groups. The CRT included 14 maternity units and data on 4503 women and was therefore well designed and of adequate power to report on the clinical outcomes chosen.

In this thesis the diagnostic cues for labour were ordered in the form of an algorithm based on the premise that structuring the judgement task (by applying a decision rule) would induce a more rational judgement process, and as a result, reduce judgement error (Hammond 1996; Hamm 1988). Other studies of decision support tools have consistently reported that they perform as well as, or better than, clinical judgement alone (Meehl 1954; Dawes et al. 2002; Grove et al. 2000; Garg et al. 2005; Kawamoto et al. 2005; Randell et al. 2007). However, relatively few studies have been conducted in real life clinical settings and, of these, most have reported on outcomes relating to process of care (for example, clinician reminder systems or diagnostic support systems which prompt specific referral pathways) rather than clinical outcomes. Kawamoto et al. (2005) reviewed RCTs of decision support systems which were tested in real world settings reporting that such systems were found to improve clinical practice in 68% of included studies. However few of these studies reported outcomes relating directly to patient care, most reported on improvements in the process of care. Similarly, Garg et al. (2005) in a review of controlled trials of computerised clinical decision support systems reported that while 64% of trials reported an improvement in clinicians' performance using computerised decision support tools, only about half of the trials included in the review reported on clinical outcomes for patients. Most of these studies were underpowered and ultimately, only seven studies reported improved clinical outcomes as a result of computerised clinical decision support. Although conducting clinical trials is challenging and results may be subject to numerous confounding factors, it is essential, if decision support systems are to

demonstrate relevance in healthcare, that they are rigorously tested in real life settings using clinical outcomes.

The fact that this is the first adequately powered study to assess the impact of diagnostic cues in early labour management is in itself a success. Health care professionals need robust evidence on which to draw in making decisions about clinical care. Yet conducting studies of complex interventions is challenging, in particular where a multi-site trial is required. The MRC framework for developing RCTs of complex interventions (MRC 2000) describes a linear process, using a mixed methods approach, in which the results of studies conducted at each step inform the next. The algorithm was developed following this framework as described in Table 1. The cues for inclusion were identified through a literature review, qualitative research on midwives' labour diagnosis and revision by experts. Once developed the algorithm was thoroughly tested using theoretical, paper based modelling and questionnaires, and found to have good face and content validity as well as a high level of inter-rater reliability. Nevertheless, this pre-clinical testing was insufficient to demonstrate whether the algorithm would be useful in a clinically relevant sense. Therefore for the final stage of testing a rigorous clinical trial methodology was used and trial outcomes were chosen which were clinically important and relevant. While the results of the CRT demonstrated that there was no statistically significant difference between the experimental and control groups for the primary outcomes, this trial has provided considerable experience of the use of the MRC framework.

The MRC describes a complex intervention as one comprising a number of components, which may act both independently and inter-dependently (MRC 2000). The framework (MRC 2000) encourages the identification of the active components of an intervention and exploration of the way in which components interact in the development stages of a trial. However, this focus, and the linear nature of the framework may result in over simplification of some aspects of the intervention and of the study design (Hawe et al. 2004). For example, in the series of studies presented in this thesis the focus was on isolating the diagnostic judgement and in developing and testing an algorithm to support that judgement. In this way a complex intervention such as a diagnostic judgement was reduced to a more simple intervention, the algorithm. While this resulted in a strong trial design (as discussed above) it may have excluded the consideration of aspects such as the different contexts into which the algorithm would be introduced. This may have contributed to the lack of difference found between experimental and control groups (Hawe et al. 2004). In the CRT a complex intervention (the algorithm) was introduced to seven maternity units which were, in themselves, complex systems. It is a characteristic of such systems that even a simple intervention may have unpredictable effects on the processes and outcomes of care (Shiell et al. 2008). It is possible that these may have contributed to the findings of no difference for many of the trial outcomes. Although useful as a way of designing robust complex or 'complicated' interventions, the MRC framework (MRC 2000) appears to have some limitations, particularly with reference to the understanding of how complex interventions may impact on the complex systems in which, in health care, they are normally introduced (Hawe et al. 2004). In conjunction with the

MRC framework, therefore, it may be useful to consider other approaches to the examination of complex interventions in health care. The science of trial development for complex interventions is constantly changing; since the inception of the trial presented in this thesis, there has been recognition of the importance of carrying out process evaluations concurrently within the trial itself (rather than in the development stages of the trial as in this study) (Oakley et al. 2006). Such an evaluation could have provided an explanation of the finding of no difference between groups for the primary outcome. However, it is also possible that the act of conducting a process evaluation during the course of a trial may in itself alter practice, thus confounding the results of the study. These issues require careful consideration during the design stages of trials of complex interventions.

In interpreting the results of a trial conducted in a clinical setting there are a number of methodological factors which must be considered, in particular, the design and power of the study, sampling and compliance with protocol.

9.1. Methodological issues

The choice of cluster randomisation for this trial was appropriate as the algorithm was targeted at the practice of midwives and individual randomisation of women or midwives could not have been used without contamination between groups. The aim was to recruit 12 maternity units with an overall target sample of 400 in each (200 before and 200 after trial implementation). Although this target was not achieved in all sites this deficit was partially offset

by the recruitment of an additional two maternity units. Units were allocated to experimental and control groups using minimisation, which is an appropriate method of allocation to group in order to maximise the balance between groups in trials such as this, where relatively few clusters are available. Group allocation was performed by the trial statistician, who was not involved in recruitment, in order to reduce potential bias due to inadequate allocation concealment (Wood et al. 2008). The balancing variable was the presence of an onsite midwife managed birth unit, chosen because midwives providing care in a birth unit would be anticipated to share a similar philosophy of care. Two units in each arm of the trial had an on-site midwife managed birth unit. The clusters in the experimental and control groups were similar in relation to size and type of maternity unit based on EGAMS classification (Scottish Executive 2002). Considering the demographic characteristics of the clusters (individuals within a cluster would be expected to be more similar to each other than to individuals in another cluster) more of the units in the control group had at least 20% of women who lived in the most deprived areas (based on the Scottish Index of Multiple Deprivation and provided by ISD specifically for this trial) although one unit in the experimental group had almost 50% of women living in the most deprived areas. However, as all but one of the eligible maternity units in Scotland participated, the sample is likely to be representative of the overall maternity population of Scotland.

The use of baseline data to reduce the in-hospital variation was a methodological development in this trial. The correlation between the

percentage of women receiving oxytocin before and after the intervention was 0.46, which was less than the 0.89 originally estimated from the data collected during the feasibility study and this reduced the power of the study (it may be that changes within the units occurred over time which reduced the correlation). Nevertheless, reflecting with hindsight on the success of this method, it can be concluded that the study had sufficient power to address the primary outcome because the 95% confidence interval for the difference in percentage use of oxytocin was –9.2 to 9.8, which excludes the difference of 10 percentage points which had been chosen as the difference which would be of clinical relevance.

It was not possible to accurately determine the number of potentially eligible women in each maternity unit and estimates were based on routinely collected data. Nor was it possible to differentiate between women who were not eligible and those who were not approached for consent to data collection. Although it appeared that in some of the smaller units almost all eligible women were included, in most of the units the proportion of eligible women not included was high and therefore selection bias could have occurred. This is a common problem in trials of intrapartum care where difficulty in estimating numbers of potentially eligible participants and high losses to recruitment are frequently reported (Hundley and Cheyne 2004). Hundley and Cheyne (2004) reviewed randomised controlled trials of intrapartum interventions in low-risk women in spontaneous labour over the period since publication of the CONSORT statement. This review found that of 15 studies identified, seven were unable to accurately identify the number of potentially eligible women. Intrapartum trials

often rely on clinical staff to seek consent from women who are in labour. This method is practical in recruiting women close to the point of study intervention, however recruitment is vulnerable both to practitioners making clinical judgements about which women to approach and to them forgetting about the trial in the midst of a busy labour suite. During the feasibility study 85% of women approached gave consent and so it is likely that the women not included in the CRT were not approached.

The strength of the cluster design is that it avoids contamination between groups, however (as with intrapartum trials) this design is reported to be prone to selection bias (Torgerson 2001) because consent to trial entry is given at cluster level, but individuals may then decide whether or not to participate in the trial intervention. The aim was that the trial would have minimum impact in the control units and this was made easier by the geographical distance between the maternity units and lack of day to day interaction between midwives across units. Further, no member of staff in control units was given access to the algorithm during the trial.

While in an ideal situation consent gaining of women in experimental and control groups would have been conducted in the same way this would not have been possible in the CRT without introducing information about the trial to labour ward midwives in the control sites, thus contaminating study groups, therefore distinct consent gaining strategies were used. In control sites no information about the trial was introduced to labour ward midwives, data collection was carried out in the postnatal wards by the local trial co-ordinators,

and the use of anonymised data was maximised in order minimise the potential for Hawthorne Effects (Braunholtz et al. 2001) and to reduce selection bias. In the experimental units midwives sought consent for data collection from women on admission for labour assessment. Although strict study entry criteria were used and all eligible women should have been approached, it is reasonable to assume that, in a real world setting, midwives will have exercised judgement in deciding who it was appropriate to approach for consent and may in particular, have been reluctant to seek consent from women who presented in advanced labour (this has been reported in other intrapartum studies (Cheyne et al. 2003)). However, systematic selection bias in the experimental group toward recruiting women admitted in early labour would be expected to have resulted in an increase in the mean time from admission to delivery in the experimental group. Although there does appear to be a small increase (table 19), this difference was not statistically significant suggesting that selection bias was not systematic. This trial brings together a research design and topic area which are recognised to be prone to selection bias and this potential must be acknowledged. However, this is compensated for by the size of the trial, the strength of the cluster design and the use of appropriate statistical techniques which control for the effects of clustering and other potential confounding factors.

9.2. Consent of midwives and use of the algorithm

Consent to study participation was sought from all midwives working in labour suites in the experimental group. The consent rate varied between units from 57% to 100%. In most (although not all) units, this consent rate reflected the

success or otherwise of subsequent data collection. The acceptability of the algorithm to midwives was assessed during the feasibility study, in which midwives reported willingness to use the algorithm. However, it is possible that this may have been the result of midwives giving socially desirable answers. The reluctance of health care professionals to use decision support has been widely reported in other studies (Corey and Merenstein 1987; Garg et al. 2005; Kawamoto et al. 2005; Thompson et al. 2004; McCaughan et al. 2005) A range of possible reasons have been suggested to explain this, for example that decision support reduces the individuality of care (Tavakoli et al. 2000; Trinder 2000), or that it undermines the skills and clinical credibility of the practitioner (Tavakoli et al. 2000; McCaughan et al. 2002; Arkes et al. 2007). In the CRT in this thesis an algorithm was completed for each woman who gave consent, with one copy being retained in the woman's case record, therefore it is clear that they were used. However, it is possible that midwives disregarded its recommendation in deciding whether to admit or discharge women. Studies of how nurses use computerised decision support tools indicate that often such tools are completed *after* the nurse has made a decision about the care that a patient should receive (O'Cathain 2004; Ruston 2006; Dowding et al. 2007).

9.3. Midwives' judgement

There is some evidence, however, that the midwives did use the algorithm and that it did alter their judgements, as significantly more women in the experimental group were discharged home following labour assessment, while women in the control group were more likely to remain in hospital from first assessment until delivery. Women in the experimental group subsequently

had significantly more pre-labour admissions, although there was no corresponding reduction in the mean length of time spent in the labour ward between final admission and delivery. It appears that women who were discharged home merely returned to the hospital creating a revolving door effect.

The results of this trial suggest that misdiagnosis of active labour is not the main reason for higher rates of intervention reported to be experienced by women who are admitted to labour wards early and that using strict labour diagnostic cues is not sufficient in itself to gate keep labour ward admissions. It appears that other factors are involved in the decision about whether to admit or discharge a woman and that these factors include the decisions made by the woman and her family. This suggestion is supported by the findings of the focus group interviews with midwives (chapter four) which identified that the labour admission assessment could be divided into the diagnostic judgement and the management decision. While the diagnostic judgement was primarily based on physical cues, the subsequent management decision (i.e. whether to admit or discharge a woman), was based on a number of additional factors relating to the institution in which care was delivered and to the woman herself. Institutional factors included pressures of workload, constraints of guidelines, the model of midwifery care and the opinion of colleagues, while factors relating to the woman included her level of distress, her expectations of labour (which midwives felt could be unrealistic), how well she was coping and the social support which was available to her at home. The midwives described having to

negotiate between these often conflicting cues and that this could lead to sub optimal care, for example giving a woman sedation, against the midwife's judgement because her mother felt she required it, or conversely, discharging a woman whom the midwife felt required care because there were no available beds.

These findings are consistent both with the suggestion that there are factors intrinsic to a woman being in the labour ward which contribute to the increased use of labour intervention (Hemminki and Simukka 1986; Holmes et al. 2001; Klein et al. 2003; Jackson et al. 2003) and also to the suggestion that there are factors intrinsic to some women which lead them to seek early admission. Studies of women's experience of early labour (Carlton et al. 2005; Cheyne et al. 2007) have found that women in their first pregnancy report feeling unprepared for the latent phase of labour and that their experience is characterised by pain and anxiety. Women seek reassurance from hospital admission and while some receive this reassurance others report feeling that their needs (in particular for pain relief) were not met. The findings in this thesis suggest that the while the algorithm has the potential to reduce admissions of women not in active labour, merely sending these women home did not produce a clinical benefit. Indeed, the findings of other studies suggest that repeated pre-labour admissions may contribute to negative childbirth experiences for women (Barnett et al. 2008).

CHAPTER 10: CONCLUSIONS

The studies reported in this thesis contribute significantly to the debate on care of women in early labour, the organisation of maternity care and maternity care research.

10.1. Care of women in early labour and the organisation of maternity care

This CRT is the first adequately powered trial of the use of explicit cues for diagnosis of active labour. The results demonstrated that the use of explicit diagnostic cues alone did not result in a reduction in oxytocin use nor in medical intervention in spontaneous labour. There was evidence that use of the algorithm did alter midwives' diagnostic judgements as significantly more women in the experimental group were discharged home following labour assessment while women in the control group were more likely to remain in hospital from first admission until they gave birth. This resulted in more pre-labour admissions for women in the experimental group while not conferring the anticipated benefits. These findings have implications for the organisation of maternity care. Current maternity service guidelines in the UK advocate advising women to remain at home or to return home until labour is established (NICE 2007) and a number of maternity units have established triage areas or telephone triage with the explicit purpose of limiting early admissions to labour wards. However the findings of the studies presented in this thesis suggest that this may be an over simplistic approach which does not address the needs of women in early labour. Women who seek hospital admission before the establishment of active labour or while in the latent phase require the skilled

care of a midwife; greater consideration of the care needs of these women is required. While some women will be happy to remain at home or to return home following labour assessment, sending women home who are not happy to do so is unlikely to be effective in reducing the rate of intervention in labour for these women.

It is possible that diagnosis and assessment of labour in the woman's own home may reduce the revolving door effect which was apparent in this CRT, in which women who were found not to be in labour and discharged merely returned to the maternity unit. The results of the Early Labour Support and Assessment (ELSA) trial are eagerly awaited (Spiby et al. 2006b). However, without a significant reorganisation of maternity services home assessment and management of early labour may not be a realistic option for the majority of women. The focus group study (chapter four) found that midwives reported having a lack of care options for women in early labour. Consideration should be given, in the design of maternity units and in the deployment of existing facilities, to providing non labour ward care for women who feel that they need the support of hospital admission while not yet in active labour.

Diagnosis of labour has been described as one of the most important and problematic judgements in the care of a woman in labour (O'Driscoll et al. 1973; Lauzon and Hodnett 2000). Despite this, few studies have investigated the way in which this key judgement may be made. The studies presented in this thesis make a significant contribution to understanding the judgement process and its

relative contribution to the overall care of a woman in labour. The findings suggest that while diagnosis of labour is an important judgement, a reduction in routine intervention in labour requires more than accurate diagnosis of labour alone. It appears that midwives may experience more difficulty with the management decision (that is, addressing the question: what should I do with this woman?) than with the initial labour diagnosis. It may be that the number of inappropriate admissions to labour wards could be reduced by supporting midwives to negotiate the complex management hurdles which accompany diagnosis of labour and by addressing the care needs of women who seek hospital admission, and require midwifery care, prior to the establishment of active labour.

10.2. Recommendations for further research

The care of women who seek hospital admission while they are not yet in labour or while in the latent phase, is an important area for further research. This thesis has focused on one specific aspect, the midwives' diagnostic judgement, and has eliminated misdiagnosis of active labour as a central cause of increased intervention in labour. Eliminating one cause however, throws the focus for further research onto other possible factors.

The model of midwives' judgement and decision-making presented in chapter four deconstructed the early labour assessment and this is valuable in signposting future research which may build on the findings of this thesis. This model identified that early labour assessment could be divided into the

diagnostic judgement (the subject of this thesis) and the management decision and that a number of factors influence the midwives' decision about whether to admit a woman in labour. These could be grouped into factors arising from the woman and those intrinsic to the institutional setting for care. Considering factors arising from the woman, further research is required on women's experiences of early labour and in particular, the factors which lead women to seek admission while they are not yet in active labour or while in the latent phase. This should include research on women's experience of pain in early labour and the management of pain in early labour as well as women's expectations of early labour, and the effect of social support before admission.

Considering the setting for care, further decision-making research is required on aspects of the current system of labour ward care which may contribute to increased intervention. For example, midwives reported that their decisions were influenced by concern about what their colleagues (in particular senior colleagues, would think of them) and that they had a role in protecting women from the effect of rigid adherence to time based protocols. Research is required to explore the effects of micro and macro decisions (that is individual and group decisions) which may lead to increased interventions in labour. In addition the interaction between these factors (the woman and her family and the maternity services) and the development of a shared decision-making intervention is a key area for further research.

10.3. The MRC Framework

The development and implementation of the series of studies presented in this thesis followed the template suggested by the MRC framework for development and evaluation of randomised controlled trials (RCTs) for complex healthcare interventions. Thus this thesis makes a valuable contribution to the experience of use of the framework in developing trials of complex healthcare interventions.

While the framework was an invaluable guide to ensuring that each aspect in the development of the intervention for the CRT was addressed and that adequate consideration was given to the methodological aspects of each stage. A contradiction was identified in the focus on isolating the active trial ingredient and standardising the trial intervention across research sites while acknowledging the interaction between components of a complex intervention and the complex systems into which it is introduced. In this thesis a complex intervention (the diagnosis of labour) was simplified to create the algorithm (the trial intervention) this was seen as both a strength and a possible weakness in the CRT design and may have contributed to the finding of no difference between groups. Future trials should give consideration to inclusion of an exploration of the way in which a trial intervention is implemented in individual research sites.

The framework advocates using a mixed method approach in which both qualitative and quantitative methods are used as appropriate, rather than

adherence to one favoured methodological approach. This means that a valuable set of data is produced at each stage which is used to develop and inform the subsequent stages. In the studies in this thesis the qualitative data provided at the development and testing stages were invaluable in interpreting the results of the CRT. Although a mixed method approach adds considerable strength to the studies presented in this thesis this approach is time consuming and labour intensive. Some contradictions in findings may occur and must be interpreted. These add to the richness of data obtained, however, it is not possible to fully explore all possible avenues of research while maintaining a focus on the ultimate aim of the research.

This research brought together both a challenging topic area and research design. Trials of intrapartum care are difficult to manage and are subject to methodological challenges in particular in relation to recruitment and adherence to protocol. Few trials in the field of midwifery care have been conducted at a national level. The study reported in this thesis was a national cluster randomised trial involving all eligible maternity units in Scotland (with the exception of one) with clinical midwives in each unit acting as local Principal Investigators and study co-ordinators. This represents a successful example of clinical academic research collaboration and contributes significantly to knowledge about the conduct of controlled trials in midwifery care.

REFERENCES

Alfirevic, Z., Devane, D., Gyte, G.M.L., 2006. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. *Cochrane Database of Systematic Reviews*, Issue 3. Art. No: CD006066. DOI: 10.1002/14651858.CD006066.

Altman, D.G.,1991. *Practical Statistics for Medical Research*. Chapman and Hall, London.

Altman, D. G., Bland, J. M., 1994. Statistics Notes: Diagnostic tests 1: sensitivity and specificity. *British Medical Journal*, 308, 1552.

Altman, D. G., Bland, J. M., 2005. Treatment allocation by minimisation. *British Medical Journal*, 330, 843.

Anderson, A.B.M., Turnbull, A.C., 1969. Relationship between length of gestation and cervical dilatation, uterine contractility, and other factors during pregnancy. *American Journal of Obstetrics and Gynecology*, December 15, 1207-1214.

Angelini, D.J., 2000. Obstetric triage and advanced practice nursing. *Journal of Perinatal and Neonatal Nursing*, 13 (4), 1-12.

Anim-Somuah, M., Smyth, R., Howell, C., 2005. Epidural versus non-epidural or no analgesia in labour. *Cochrane Database of Systematic Reviews*, Issue 4. Art. No.: CD000331. DOI: 10.1002/14651858.CD000331.pub2.

Arkes, H.R., Shaffer, V.A., Mitchell, A.M., 2007. Patients derogate physicians who use a computer-assisted diagnostic aid. *Medical Decision-making,* 27, 189-202.

Austin, D.A., Calderon, L., 1999. Triaging patients in the latent phase of labour. *Journal of Nurse Midwifery*, 44 (6), 585 – 91.

Ball, J.A., Washbrook, M., 1996. *Birthrate Plus: A framework for workforce planning and decision-making for midwifery services*. London: Books for Midwives Press.

Ballantyne, J. W., Browne, F.J., 1922. The problems of foetal post-maturity and prolongation of pregnancy. *The Journal of Obstetrics and Gynaecology of the British Empire,* 29 (2), 177- 237.

Barnett, C., Hundley, V., Cheyne, H., Kane, F., 2008. 'Not in labour': impact of sending women home in the latent phase. *British Journal of Midwifery*, 16 (3). 144-153.

Baron, J., 2000. *Thinking and Deciding*. 3rd ed. Cambridge: Cambridge University Press.

Barrows, H.S., Pickell, G.C., 1991. Developing clinical problem-solving skills: A guide to more effective diagnosis and treatment. New York : WW Norton & Company.

Bekker, H., Thornton, J.G., Airey, C.M., Connelly, J.B., Hewison, J., Robinson, M.B., Lilleyman, J., MacIntosh, M., Maule, A.J., Michie, S., Pearman, A.D., 1999. Informed decision-making: an annotated bibliography and systematic review. Health Technology Assessment, 3 (1), 1-156.

Benner, P., 1982. From novice to expert. American Journal of Nursing, 82 (3), 402-407.

Benner, P., Tanner, C., 1987. Clinical judgement: how expert nurses use intuition. *American Journal of Nursing*, 87 (1), 23-31.

Berman, D.A., Coleridge, S.T., McMurry, T.A., 1989. Computerized algorithm directed triage in the emergency department. *Annals of Emergency Medicine*, 18 (2), 141-144.

Bernal, A.L., 2001. Overview of current research in parturition. *Experimental Physiology*, 86 (2), 213-222.

Bick, D., 2006. A reflection on the last 21 years; the coming of age of "Midwifery". *Midwifery*, 22 (1), 1-2.

Bishop, E.H., 1964. Pelvic scoring for elective induction. *Obstetrics and Gynecology*, 24, 266-268.

Bland, J. M., 1997. Statistical notes. Trials randomised in clusters. *British Medical Journal*, 315 (7018), 600.

Bland, M., 2005. *Measuring health and disease: Cohen's Kappa*. York. University of York Department of Health Sciences. Available from: http://www-users.york.ac.uk/~mb55/

Bless, H., Clore, G.L., Schwarz, N., Golisano, V., Rabe, C., Wolk, M., 1996. Mood and the use of scripts: Does a happy mood really lead to mindlessness? *Journal of Personality and Social Psychology*, 71 (4), 665-679.

Bowling, A., 1991. *Measuring disease*. Open University Press, Buckingham.

Boylan, P., Frankowski, R., Rountree, R., Selwyn, B., Parrish, K., 1991. Effect of active management of labour on the incidence of cesarean section for dystocia in nulliparas. *American Journal of Perinatology*, 8 (6), 373-379.

Bramwell, R., West, H., Salmon, P., 2006. Health professionals' and service users' interpretation of screening test results: experimental study. *British Medical Journal,* 333 (7562), 284. doi:10.1136/bjm.38884.663102.AE

Braunholtz, D.A., Edwards, S.J., Lilford, R.J., 2001. Are randomized clinical trials good for us (in the short term)? Evidence for a "trial effect". *Journal of Clinical Epidemiology*, 54, 217-224.

Brill, S., Gurman, G.M., Fisher, A., 2003. A history of neuraxial administration of local analgesics and opioids. *European Journal of Anaesthesiology*, 20, 682-689.

Brillman, J.C., Dozema, D., Tandberg, D., Sklar, D.P., Davis, D.P.,Simms, S., Skipper, B.J., 1996. Triage: limitations in predicting need for emergent care and hospital admission. *Annals of Emergency Medicine*. 27 (4), 493-500.

Brunswik, E., 1952. The conceptual framework of psychology. In: *International encyclopaedia of unified science*. USA: University of Chicago Press. vol 1, no 10, p 4-102.

Burvill, S., 2002. Midwifery diagnosis of labour onset. *British Journal of Midwifery,* 10 (10), 600- 605.

Cairns, J., Shackley, P., Hundley, V., 1996. Decision-making with respect to diagnostic testing. *Medical decision-making*, 16 (2), 161-168.

Callister, L.C., Khalaf, I., Semenic, S., Kartchner, R., Vehvilainen-Jalkunen, K., 2003. The pain of childbirth: perceptions of culturally diverse women. *Pain Management in Nursing*, 4 (4), 145-154.

Campbell, H., Hotchkiss, R., Bradshaw, N., Porteous, M., 1998. Integrated care pathways. *British Medical Journal*. 316 (7125), 133 -137.

Campbell, M.K., Elbourne, D.R., Altman, D.G., 2004. CONSORT statement: extension to cluster randomised trials. *British Medical Journal*, 328. 702-708.

Campbell, M.K., Mollison, J., Steen, N., Grimshaw, J.M., Eccles, M., 2000. Analysis of cluster randomized trials in primary care: a practical approach. *Family Practice*, 17 (2), 192-196.

Campbell, R., Macfarlane, A., 1994. *Where to be born? The debate and the evidence.* 2nd Edition*.* Oxford: National Perinatal Epidemiology Unit.

Carlton, T., Callister, L.C., Stoneman, E., 2005. Decision-making in laboring women: ethical issues for perinatal nurses. *Journal of Perinatal and Neonatal Nursing*, 19 (2), 145-154.

Cartmill, R.S.V., Thornton, J.C., 1992. Obstetric decision-making: the effect of varying the presentation of partogram information. *Lancet*, 339, 1520 -1522.

CEMACH, 2007. *Saving mothers' lives: reviewing maternal deaths to make motherhood safer – 2003-2005. The seventh report of the confidential enquiries into maternal deaths in the United Kingdom.* London: CEMAC. Available from: www.cemach.org.uk.

Challis, J.R.G., Mathews, S.G., Gibb, W., Lye, S.J., 2000. Endocrine and paracrine regulation of birth at term and preterm. *Endocrine Reviews*, 21 (5), 514-550.

Chalmers, B., 1992. WHO appropriate technology for birth revisited. *British Journal of Obstetrics and Gynaecology,* 99 (9), 709-710.

Cheyne, H., Dowding, D., Hundley, V., Aucott, L., Styles, M,. Mollison, J., Greer, I., Niven, C., 2008. The development and testing of an algorithm for diagnosis of active labour in primiparous women. *Midwifery*, 24 (2), 199-213. *(Epub* 2007).

Cheyne, H., Dunlop, A., Shields, N., Mathers, A., 2003. A randomised controlled trial of admission electronic fetal monitoring in normal labour. *Midwifery*, 19 (3), 221-229.

Cheyne, H., Terry, R., Niven, C., Dowding, D., Hundley, V., McNamee, P., 2007. 'Should I come in now?' a study of women's early labour experiences. *British Journal of Midwifery*, 15 (10), 604-609.

Cioffi, J., Markham, R., 1997. Clinical decision-making by midwives; managing case complexity. *Journal of Advanced Nursing,* 25 (2), 265-272.

Cioffi, J., 1998. Decision-making by emergency nurses in triage assessments. *Accident and Emergency Nursing*, 6, 184-191.

Considine, J., Ung, L., Thomas, S., 2000. Triage nurses' decisions using the National Triage Scale for Australian emergency departments. *Accident and Emergency Nursing,* 8, 201-209.

Corey, G.A., Merenstein, J.H., 1987. Applying the acute ischemic heart disease predictive instrument. *Journal of Family Practice*, 25, 127-133.

Costello, A., Osrin, D., 2005. Epidemiological transition, medicalisation of childbirth, and neonatal mortality: three Brazilian birth-cohorts. *Lancet*, 365 (9462), 825-826.

Creedy, D.K., Shochet, I.M., Horsfall, J., 2000. Childbirth and the development of acute trauma symptoms: Incidence and contributing factors. *Birth*, 27, 104-111.

Crowther, C., Enkin, M., Keirse, M.J.N.C., Brown, I., 1991. Monitoring the progress of labour. *In:* Chalmers, I., Enkin, M., Keirse, M.J.N.C., eds. *Effective care in pregnancy and childbirth*. Oxford: Oxford University Press, 833-845.

Csapo, A.I., 1961. The onset of labour. *Lancet*, 2 (7197), 277-80.

Dalgleish, L.I., de Michele, E., 1995*. Risk assessment and decision-making in child protection.* Presented at the 5th Australian conference on Child Abuse and Neglect. Melbourne.

Davis-Floyd, R.E., Davis, E., 1997. Intuition as authoritative knowledge in midwifery and home birth. In: Davis –Floyd, R.E., Sargent, C.F., eds. *Childbirth and authoritative knowledge: cross cultural perspectives*. Berkley: University of California Press. 315-349.

Dawes, R.M., Faust, D., Meehl, P.E., 2002. Clinical versus actuarial judgement. In: Gilovich, T., Griffin, D., Kahneman, D., eds. *Heuristics and Biases: the psychology of intuitive judgement*. Cambridge: Cambridge University Press. 716-729.

Department of Health, 1998. *NHS Maternity Statistics, England: 1989-90 to 1994-95*. London: Department of Health. Available from: http://www.dh.gov.uk/en/Publicationsandstatistics/Statistics/DH_4010122

Department of Health, 2005. *NHS maternity statistics, England: 2003-04. Statistical Bulletin 2005/10.* London: Department of Health. Available from: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsStatistics/DH_4107060

den Hertog, C.E., de Groot, A.N., van Dongen, P.W., 2001. History and use of oxytocics. *European Journal of Obstetrics Gynecology and Reproductive Biology,* 94 (1), 8-12.

Dhami, M.K., Harries, C., 2001. Fast and frugal versus regression models of human judgement. *Thinking and reasoning,* 7 (1), 5-27.

Donner, A., 1998. Some aspects of the design and analysis of cluster randomised trials. *Applied Statistics*, 47, (part 1) 95-113.

Donner, A., Klar, N., 2000. *Design and analysis of cluster randomization trials in health research*. London: Arnold.

Dowding, D., Lattimer, V., Randell, R., Mitchell, N., Foster, B., Thompson, C., Cullum, N., Webster, A., Summers, R., Lathlean, J., Crouch, R., Owens, R., 2007. *How do nurses use new technologies to inform their decision-making?* Final Report to the policy research programme, Department of Health. York: University of York.

Dowding, D., Thompson, C., 2002. Decision analysis. In: Thompson, C., Dowding, D., eds. *Clinical Decision-making and Judgement in Nursing*. China: Churchill Livingstone, 131-145.

Dowding, D., Thompson, C., 2003. Measuring the quality of judgement and decision-making in nursing. *Journal of Advanced Nursing*, 44(1), 49-57.

Dowie, J., 1996. The research-practice gap and the role of decision analysis in closing it. *Health Care Analysis*, 4, 5-18.

Dowie, J., 1993. Clinical decision analysis: Background and introduction. In: Llewelyn, H., and Hopkins, A., eds. *Analysing how we reach clinical decisions.* London: Royal College of Physicians. 7-26.

Downe, S., 2001. Is there a future for normal birth? *The Practicing Midwife*, 4 (6), 10-12.

Downe, S., 2004. *Normal Childbirth: evidence and debate*. China: Churchill Livingstone.

Dreyfus, H.L., Dreyfus, S.E., 1986. *Mind over machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: The Free Press.

Edwards, S.J.L., Braunholtz, D.A., Lilford, R.J., Stevens, A.J., 1999. Ethical issues in the design and conduct of cluster randomised controlled trials. *British Medical Journal,* 318, 1407 – 1409.

Edwards, W., 1954. The theory of decision-making. *Psychological Bulletin,* 51, 380-417.

Edwards, W., Lindman, H., Savage, L.J., 1963. Bayesian statistical inference for psychological research. *Psychological Review*, 70, 193-242.

Edwards, W., Newman, R.J., 2000. Multiattribute Evaluation. In: Connolly, T., Arkes, H.R., Hammond, K.R., eds. *Judgement and Decision-making: an interdisciplinary reader.* 2nd ed. Cambridge: Cambridge University Press. 17-35.

Elbourne D.R., Campbell, M. K., 2001. Extending the CONSORT statement to cluster randomised trials: for discussion. *Statistics in Medicine*, 20, 489-496.

Elstein, A.S., Schwarz, A., 2002. Clinical problem solving and diagnostic decision-making: selective review of the cognitive literature. *British Medical Journal*, 324, 729-732.

Elstein, A., Shulman, L.S., Sprafka, S.A., 1990. Medical problem solving: a ten-year retrospective. *Evaluation and the Health Professions*, 13 (1), 5-36.

English, I., 1993. Intuition as a function of the expert nurse: a critique of Benner's novice to expert model. *Journal of Advanced Nursing,* 18, 387-393.

Finucane, M.L., Alhakami, A., Slovic, P., Johnson, S.M., 2000. The affect heuristic in judgements of risks and benefits. *Journal of Behavioural Decision-making*, 13, 1-17.

Flaskerud, J.H., 1979. Use of vignettes to elicit responses to broad concepts. *Nursing Research*, 28, 210-2.

Fleiss, J.L., 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76. 378-382.

Freeman, J.M., 2007. Beware: the misuse of technology and the law of unintended consequences. *Neurotherapeutics*, 4 (3), 549-554.

Friedman, E., 1989. Normal and dysfunctional labour. *In*: Cohen, W.R., Acker, D.B., Friedman, E.A., eds. *Management of Labour*. 2nd Ed. Rockville, MD: Aspen. 1-18

Frigoletto, F.D., Lieberman, E., Lang, J.M., Cohen, A., Barss, V., Ringer, S., Datta, S., 1995. A clinical trial of active management of labour. *New England Journal of Medicine,* 333 (12), 745-750.

Frisch, D., Clemen, R.T., 1994. Beyond expected utility; rethinking behavioural decision research. *Psychological Bulletin,* 116, 48.

Fuchs, A.R., Fuchs, F., 1984. Endocrinology of human parturition: a review. *British Journal of Obstetrics and Gynaecology*, 91, 948-967.

Garcia, J., Garforth, S., 1989. Labour and delivery routines in English consultant maternity units. *Midwifery,* 5 (4), 155-162.

Garfield, R.E., Sims, S., Daniel, E.E., 1977. Gap junctions: their presence and necessity in myometrium during gestation. *Science*, 98, 958-960.

Garfield, R.E., Merrett, D., Grover, A.K., 1980. Gap junction formation and regulation in myometrium. *American Journal of Physiology*, 239, C217-C228.

Garg, A.X., Adhikari, N.K.J., McDonald, H., Rosas-Arellano, M.P., Devereaux, P.J., Beyene, J., Sam, J., Haynes, R.B., 2005. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. A systematic review. *Journal of the American Medical Association*, 293(10), 1223-1238.

Gaskin, I.M., 2002. *Spiritual Midwifery*. Cambridge: Summertown.

Gigerenzer, G., 1991. How to make cognitive illusions disappear: Beyond "Heuristics and Biases". *European Review of Social Psychology*, 2, 83-115.

Gigerenzer, G., 2002. *Reckoning with risk: learning to live with uncertainty*. London: Penguin.

Gilovich, T., Griffin, D., 2002. Introduction – Heuristics and Biases: then and now. In: Gilovich, T., Griffin, D., Kahneman, D., eds. 2002. *Heuristics and Biases: the psychology of intuitive judgement*. Cambridge: Cambridge University Press. 1-18.

Goldberg, L.R., 1968. Simple models or simple processes? Some research on clinical judgement. *American Psychologist,* 23, 483-496.

Goldberg, L.R., 1970. Man versus model of man: A rationale, plus some evidence for a method of improving on clinical inferences. *Psychological Bulletin,* 73, 422-432.

Gould, D., 1996. Using vignettes to collect data for nursing research studies: how valid are the findings? *Journal of Clinical Nursing*, 5 (4), 207-12.

Gould, D., 2000. Normal labour: a concept analysis. *Journal of Advance Nursing,* 31 (2), 418-427.

Grant, A.M., Altman, D.G., Babiker, A.B., Campbell, M.K., Clemens, F.J., Darbyshire, J.H., Elbourne, D.R., McLeer, S.K., Parmar, M.K., Pocock, S.J., Spiegelhalter, D.J., Sydes, M.R., Walker, A.E., Wallace, S.A., DAMOCLES study group. 2005. Issues in data monitoring and interim analysis of trials. *Health Technology Assessment,* 9 (7), 1-238, iii-iv.

Green, J. M., Baston, H. A., 2007. Have women become more willing to accept obstetric interventions and does this relate to mode of birth? Data from a prospective study. *Birth,* 34, 6-13.

Greer, I.A., 1995. The biochemistry and physiology of labour. In: Chamberlain, G.V.P., ed. *Turnbulls Obstetrics*. London: Churchill Livingstone. 551-567.

Grove, M., Zald, D.H., Lebow, B.S., Snitz, B.E., Nelson, C., 2000. Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12 (1), 19-30.

Gülmezoglu, A.M., Crowther, C.A., Middleton, P., 2006. Induction of labour for improving birth outcomes for women at or beyond term. *Cochrane Database of Systematic Reviews*, Issue 4. Art. No.: CD004945. DOI: 10.1002/14651858.CD004945.pub2.

Hacking, A., 1990. *The taming of chance*. Cambridge: Cambridge University Press. 150-160.

Hamm, R.M., 1988. Clinical Intuition and Clinical Analysis: Expertise and the Cognitive Continuum. In: Dowie, J., Elstein, A., eds. *Professional Judgement: a reader in clinical judgement and decision-making.* Cambridge: Cambridge University Press. 78-109.

Hammond, K.R., 1996. *Human Judgement and Social Policy: irreducible uncertainty, inevitable error, unavoidable injustice*. Oxford: Oxford University Press.

Hansen, A.K., Wisborg, K., Uldbjerg, N., Henriksen, T.B., 2008. Risk of respiratory morbidity in term infants delivered by elective caesarean section: a cohort study. *British Medical Journal*, 336 (7635), 85-87.

Hastie, R., Dawes, R.M., 2001. *Rational Choice in an uncertain world: the psychology of judgement and decision-making*. California: Sage Publications.

Hawe, P., Sheill, A., Riley, T. 2004. Complex interventions: how “out of control” can a randomised controlled trial be? *British Journal of Medicine*, 328, 1561-1563.

Hayward, A.C., Harling, R., Wetten, S., Johnson, A.M., Munro, S., Smedley, J., Murad, S., Watson, J.M., 2006. Effectiveness of an influenza vaccine programme for care home staff to prevent death, morbidity, and health service use among residents: cluster randomised controlled trial. *British Medical Journal,* 333 (7581), 1241. Epub 2006 Dec 1.

Hemminki, E., Simukka, R., 1986. The timing of hospital admission and progress of labour. *European journal of obstetrics, gynecology, and reproductive biology,* 22 (1-2), 85-94.

Hollis, B., 2002. Prolonged pregnancy. *Current Opinion in Obstetrics and Gynecology*, 14, 203 – 207.

Holmes, P., Oppenheimer, L.W., Wen, S.W., 2001. The relationship between cervical dilatation at initial presentation in labour and subsequent intervention. *British Journal of Obstetrics and Gynaecology,* 108, 1120-1124.

Hughes, R., Huby, M., 2002. The application of vignettes in social and nursing research. *Journal of Advanced Nursing*, 37(4), 382-6.

Hundley, V., Cheyne, H., 2004. The trials and tribulations of intrapartum studies. *Midwifery*, 20, 27-36.

Hundley, V., Cruickshank, F.M., Lang, G.D., Glazener, C.M., Milne, J.M., Turner, M., Blyth, D., Mollison, J., Donaldson, C., 1994. Midwife managed delivery unit: a randomised controlled comparison with consultant led care. *British Medical Journal*, 309 (6966), 1400-1404.

Hunter, B., 2007. *The All Wales Clinical Pathway for Normal Labour: What are the experiences of midwives, doctors, managers and mothers?* Wales: Swansea University.

Hutton, J.L., 2001. Are distinctive ethical principles required for cluster randomized trials? *Statistics in Medicine*, 20, 473-488.

Inch, S., 1985. Management of the third stage of labour – another cascade of intervention? *Midwifery*, 1 (2), 114 -122.

Information and Statistics Division (ISD), 1999. *Scottish Health Statistics: Births & Babies*. Edinburgh: ISD.

Information and Statistics Division (ISD), 2006. *Scottish Health Statistics: Births & Babies*. Edinburgh: ISD. Available from: http://www.isdscotland.org

Jackson, D.J., Lang, J.M., Ecker, J., Swartz, W.H., Heeren, T., 2003. Impact of collaborative management and early admission in labour on method of delivery. *Journal of Obstetric, Gynecologic, and Neonatal Nursing*, 32 (2), 147-157.

Jackson, P., 1998. Focus group interviews as a methodology. *Nurse Researcher,* 6 (1), 72-84.

Janssen, P., Iker, C., Carty, E., 2003. Early labour assessment and support at home: a randomised controlled trial. *Journal of Obstetrics and Gynaecology Canada,* 25 (9), 734-741.

Johanson, R., Newburn, M., Macfarlane, A., 2002. Has the medicalisation of childbirth gone too far? *British Medical Journal*, 324, 892-895.

Kahneman, D., Frederick, S., 2002. Representative revisited : attribute substitution in intuitive judgement. In: Gilovich, T., Griffin, D., Kahneman, D., eds. *Heuristics and Biases; the psychology of Intuitive Judgement*. Cambridge*:* Cambridge University Press. 49-81.

Kahneman, D., Frederick, S., 2005. A Model of Heuristic Judgement. In: Holyoak, K.J., Morrison, R.G., eds. *The Cambridge Handbook of Thinking and Reasoning*. 267-293.

Kahneman, D., Slovic, P., Tversky, A., 1982. *Judgement under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press.

Kortenoever, M.E., 1950. Pathology of pregnancy: Pregnancy of long duration and post mature infant. *Obstetrical and Gynecological Survey*, 5, 812- 814.

Kawamoto, K., Houlihan, C.A., Balas, E.A., Lobach, D.F., 2005. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *British Medical Journal*, 330 (7494), 765-768.

Keelan, J.A., Coleman, M., Mitchell, M.D., 1997. The molecular mechanisms of term and preterm labour: recent progress and clinical implications. *Clinical Obstetrics and Gynecology,* 40 (3), 460 – 478.

Kelly, M.A., 1999. Triage and management of the pregnant hypertensive patient. *Journal of Nurse Midwifery*, 44 (6), 558-571.

Kerry, S.M., Bland, J.M., 1998. Analysis of trials randomised in clusters. *British Medical Journal,* 316. 54.

Kitzinger, S., 2006. *Birth Crisis*. Oxon: Routledge.

Klein, M.C., 2006. Epidural analgesia: Does it or doesn't it? *Birth*, 33, 74-76.

Klein, M.C., Kelly, A., Kaczorowski, J., Grzybowski, S., 2003. The effect of family physician timing of maternal admission on proceedures in labour and maternal and infant morbidity. *Journal of Obstetrics and Gynecology Canada,* 26 (7), 641-645.

Klein, M.C., Sakala, C., Simkin, P., Davis- Floyd, R., Rooks, J.P., Pincus, J., 2006. Why do women go along with this stuff? Roundtable discussion: Part 2. *Birth,* 33, 245-250.

Knight, M., Kurinczuk, J.J., Spark, P., Brocklehurst, P., United Kingdom Obstetric Surveillance System Steering Committee., 2008. Cesarean delivery and peripartum hysterectomy. *Obstetrics and Gynecology*, 111 (1), 97-105.

Lane, P., McKenna, H., Ryan, A., Fleming, P. (2001). Focus group methodology. *Nurse Researcher,* 8 (3), 45-49.

Lauzon, L., Hodnett, E., 2000. Caregivers' use of strict criteria for diagnosing active labour in term pregnancy (Cochrane Review). In: *The Cochrane Library*, Issue 4. Oxford: Update Software.

Lauzon L, Hodnett E. Labour assessment programs to delay admission to labour wards. *Cochrane Database of Systematic Reviews* 2001, Issue 3. Art. No.: CD000936. DOI: 10.1002/14651858.CD000936.

Leli, D.A., Filskov, S.B., 1984. Clinical detection of intellectual deterioration associated with brain damage. *Journal of Clinical Psychology,* 40, 1435-1441.

Letourneau, S.; Jensen, L., 1998. Impact of a decision tree on chronic wound care. *Journal of Wound, Ostomy and Continence Nursing*, 25 (5): 240-247.

Lopez-Zeno, J.A., Peaceman, A.M., Adashek, J.A., Socol, M.L., 1992. A controlled trial of a partogram for the active management of labour. *New England Journal of Medicine,* 326 (7), 450-454.

Luby, S.P., Agboatwalla, M., Painter, J., Altaf, A., Billhimer, W., Keswick, B., Hoekstra, R.M., 2006. Combining drinking water treatment and hand washing for diarrhoea prevention, a cluster randomised controlled trial. *Tropical Medicine and International Health,* 11 (4), 479 -489.

Ludwick, R., Zeller, R.A., 2001. The factorial Survey: An experimental Method to Replicate Real Problems. *Nursing Research,* 50 (2), 129-133.

MacArthur, C., Glazener, C.M., Wilson, P.D., Herbison, G.P., Gee, H., Lang, G.D., Lancashire, R., 2001. Obstetric practice and faecal incontinence three months after delivery. *British Journal of Obstetrics and Gynaecology,* 108 (7), 678 -683.

MacArthur, C., Winter, H.R., Bick, D.E., Knowles, H., Lilford, R., Henderson, C., Lancashire, R.J., Braunholtz, D.A., Gee, H., 2002. Effects of redesigned community postnatal care on womens' health 4 months after birth: a cluster randomised trial. *Lancet*, 359 (9304), 378- 385.

Macfarlane, A., Mugford, M., 2000. *Birth counts: statistics of pregnancy and childbirth,* vol 1 and 2. London, HMSO.

Machin, D., Scamell, M., 1997. The experience of labour: using ethnography to explore the irresistible nature of the bio-medical metaphor during labour. *Midwifery*, 13, 78 – 84.

Mahlmeister, L., van Mullem, C., 2000. The process of triage in perinatal settings: clinical and legal issues. *Journal of Perinatal Neonatal Nursing*, 13, 13-30.

Mander, R., 1998. A reappraisal of Simpson's introduction of chloroform. *Midwifery,* 14 (3), 181-189.

Martin, C.B. Jr., 1998. Electronic fetal monitoring: a brief summary of its development, problems and prospects. *European Journal of Obstetrics Gynecology and Reproductive Biology*, 78 (2), 133-140.

McCarney, R., Warner, J., Iliffe, S., van Haselen, R., Griffen, M., Fisher, P., 2007. The Hawthorne Effect: a randomised controlled trial. *BMC Medical Research Methodology* 7, 30. doi: 10.1186/1471-2288-7-30.

McCaughan, D., Thompson, C., Cullum, N., Sheldon, T., Thompson, D.R., 2002. Acute care nurses' perceptions of barriers to using research information in clinical decision-making. *Journal of Advanced Nursing*, 39 (1), 46-60.

McCaughan, D., Thompson, C., Cullum, N., Sheldon, T., Raynor, P., 2005. Nurse practitioner and practice nurses' use of research information in clinical decision-making: findings from an exploratory study. *Family Practice* 22 (5), 490-497.

McClean, J., Turnbull, D.A., McGinley, M.C., Lunan, C.B., 1999. Can childbirth be changed? Offering women a choice. *British Journal of Midwifery*. 7 (10) 622-626.

McLean, M., 2001. Timing the onset of parturition. *Frontiers of Hormone Research,* 27, 229-245.

McNiven, P.S., Williams, J.I., Hodnett, E., Kaufman, K., Hannah, M.E., 1998. An early labor assessment program: a randomized, controlled trial. *Birth,* 25 (1), 5-10.

Meehl, P.E., 1954. *Clinical versus statistical prediction: A theoretical analysis and review of the evidence*. Minneapolis: University of Minnesota Press.

Medical Research Council, 2000. *A framework for development and evaluation of RCTs for complex interventions to improve health*. Medical Research Council. Available from: http://www.mrc.ac.uk

Medical Research Council, 2002. *Cluster randomised trials: methodological and ethical considerations.* MRC clinical trials series. Medical Research Council. Available from: http://www.mrc.ac.uk

Midwifery Development Unit, 1995. *The establishment of a midwifery development unit at Glasgow Royal Maternity Hospital*. Glasgow: Midwifery Development Unit.

Mold, J.W., Stein, H.F., 1986. The cascade effect in the clinical care of patients. *New England Journal of Medicine,* 314 (1986), 512–524.

Mollison, J., Simpson, .JA., Campbell, M.K., Grimshaw, J.M., 2000. Comparison of analytical methods for cluster randomised trials: and example from a primary care setting. *Journal of Epidemiology and Biostatistics,* 5 (6), 339-348.

Morgan, D.L., 1995. Why things (sometimes) go wrong in focus groups. *Qualitative Health Research*, 5 (4), 516-523.

Morse, J.M., Field, P.A., 1996. *Nursing Research: The application of qualitative approaches* 2nd ed. London: Chapman & Hall.

Murphy-Black, T., 1992. Systems of midwifery care in use in Scotland. *Midwifery*, 8 (3), 113-24.

Murray, S., 2000. Relation between private health insurance and high rates of caesarean section in Chile: qualitative and quantitative study. *British Medical Journal,* 321, 1501-1505.

Nathanielsz, P.W., 1994. A time to be born: implications of animal studies in maternal-fetal medicine. *Birth*, 21 (3), 163-169.

National Collaborating Centre for Women's and Children's Health. 2004. Caesarean section: clinical guideline. London: RCOG Press. Available from: www.nice.org.uk/pdf/CG013NICEguideline.pdf; http://www.nice.org.uk/pdf/CG013fullguideline.pdf.

National Collaborating Centre for Women's and Children's Health. 2007. *Intrapartum care: care of healthy women and their babies during childbirth*. Clinical guideline. London: RCOG Press. Available from: http://guidance.nice.org.uk/CG55/niceguidance/pdf.

National Library of Medicine, 1998. *Cesarean Section – a brief history*. Available from: http://nlm.nih.gov/exhibition/cesarean/cesarean_2.html

NCT/RCM/RCOG, 2007. Making normal birth a reality: Consensus statement from the Maternity Care Working Party. NCT/RCM/RCOG. Available from: www.rcog.org.uk/resources/public/pdf/**normal_birth_consensus**.pdf.

NHS Wales, 2006. *All Wales Clinical Pathway for Normal Labour.* Available from: www.**wales**.nhs.uk/**normallabour**/

Niven, C.A., 1993. *Psychological care for families: Before, during and after birth.* Oxford: Butterworth Heinemann.

Oakley, A., Strange, V., Bonell, C., Allen, E., Stephenson, J. 2006. Process evaluation in randomised controlled trials of complex interventions. *British Medical Journal*, 332, 413-416.

O'Cathain, A., Sampson, F.C., Munro, J.F., Thomas, K.J., Nicholl, J., 2004. Nurses' views of using computerized decision support software in NHS Direct. *Journal of Advanced Nursing*, 45*,* 280-286*.*

O'Connor, A., Jacobsen, M.J., 2003. *Workbook on developing and evaluating patient decision aids*. Available from: http://decisionaid.ohri.ca/docs/develop/Develop_DA.pdf.

O'Driscoll, K., Foley, M., MacDonald, D., 1984. Active management of labor as an alternative to Cesarean Section for Dystocia. *Obstetrics and Gynecology,* 63 (4), 485-490.

O'Driscoll, K., Meagher, D., 1980. *Active management of labour*. London: Saunders.

O'Driscoll, K., Stronge, J.M., Minogue, M., 1973. Active Management of Labour. *British Medical Journal,* 3 (5872)**,** 135-137.

Paley, J., 2006 Evidence and expertise. *Nursing Inquiry,* 13 (2), 82-93.

Paley, J., Cheyne, H., Dalgleish, L., Duncan, E.A., Niven, C.A., 2007. Nursing's ways of knowing and dual process theories of cognition. *Journal of Advanced Nursing*, 60 (6), 692-701.

Patel, V.L., Kaufman, D.R., Arocha, J.F., 2002. Emerging paradigms of cognition in medical decision-making. *Journal of Biomedical Informatics,* 35, 52-75.

Petrou, S., Glazener, C., 2002. The economic costs of alternative models of delivery during the first two months postpartum: results from a Scottish observational study. *British Journal of Obstetrics and Gynaecology*, 109, 214-217.

Potter, J.E., Hopkins, K., Faundes, A., Perpetuo, I., 2008. Women's autonomy and scheduled caesarean sections in Brazil: A cautionary tale. *Birth*, 35, 33-40.

Rahnama, P., Ziaei, S., Faghidzadeh, S., 2006. Impact of early admission in labour on method of delivery. *International Journal of Gynecology and Obstetrics*, 92, 217-220.

Randell, R., Mitchell, N., Dowding, D., Cullum, N., Thompson, C., 2007. Effects of computerized decision support systems on nursing performance and patient outcomes: a systematic review. *Journal of Health Services Research and Policy,* 12 (4), 242-249.

Redshaw, M., Rowe, R., Hockley, C., Brocklehurst, P., 2007. *Recorded delivery: a national survey of women's experience of maternity care 2006.* Oxford: National Perinatal Epidemiology Unit. Available at: http://www.npeu.ox.ac.uk/downloads/maternitysurveys/maternity-survey-report.pdf

Roberts, C.L., Tracy, S,. Peat, B., 2000. Rates for obstetric intervention among private and public patients in Australia: population based descriptive study. *British Medical Journal,* 321, 137-141.

Robinson, N., 1999. The use of focus group methodology – with selected examples from sexual health research. *Journal of Advanced Nursing*, 29 (4), 905-913.

Rosenthal, G.E., Mettler, G., Pare, S., Riegger, M., Ward, M., Landefeld, C.S., 1992. Diagnostic judgements of nurse practitioners providing primary gynaecologic care: a quantitative analysis. *Journal of General Internal Medicine,* 7 (3), 304-311.

Royal College of Obstetricians and Gynaecologists, 2001. National evidence based clinical guidelines. *The use of Electronic Fetal Monitoring*. London: RCOG. Available from: www.**rcog**.org.uk.

Royal College of Midwives, 2005. *Evidence based guidelines for midwifery-led care in labour.* Midwifery Practice Guideline No: 1. Available from: http://www.rcm.org.uk.

Ruston, A., 2006. Interpreting and managing risk in a machine bureaucracy: Professional decision-making in NHS Direct*.* *Health, Risk and Society,* 8 (3), *257-271.*

Ryan, M., Scott, D.A., Reeves, C., Bate, A., van Teijlingen, E.R., Russell, E.M., Napper, M., Robb, C.M., 2001. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technology Assessment,* Vol 5: No 5. 1-186.

Sadler, L.C., Davison, T., McCowan, L.M., 2000. A randomised controlled trial and meta-analysis of active management of labour. *British Journal of Obstetrics and Gynaecology*, 107 (7), 909-915.

Sandall, J., 2004. Promoting normal birth: weighing the evidence. In: Downe, S., ed. *Normal Childbirth: evidence and debate*. China: Churchill Livingstone.

Scottish Executive, 2002. *Report of the expert group on acute maternity services (EGAMS)*. Edinburgh: Scottish Executive.

Scottish Programme for Clinical Effectiveness in Reproductive Health (SPCERH), 2001. *Scottish audit of the prevention and management of emergencies in labour: an audit of progress towards safer childbirth.* Aberdeen: Dugald Baird Centre.

Scottish Programme for Clinical Effectiveness in Reproductive Health (SPCERH), 2003. *Operative vaginal delivery in Scotland, a 20 year overview*. Scotland ISD. Available from: www.**isdscotland**.org.

Segal, D., Sheiner, E., Yohai, D., Shoham-Vardi, I., Katz, M., 1999. Early amniotomy – high risk factor for cesarean section. *European Journal of Obstetrics and Gynecology and Reproductive Biology*. 86 (2), 145-149.

Shiell, A,. Hawe, P., Gold, L. 2008. Complex interventions or complex systems? Implications for health economic evaluation. *British Medical Journal,* 336, 1281-1283.

Simon, H.A., 1983. *Reason in human affairs*. Stanford, CA: Stanford University Press p 13-14.

Simonda, W., 2002. Watching the clock: keeping time during pregnancy, birth, and postpartum experiences. *Social Science and Medicine,* 55 (4), 559-570.

Sloman, S.A., 1996. The empirical case for two systems of reasoning. *Psychological Bulletin,* 119 (1), 3-22.

Smith, G.C.S., 2001. Use of time to event analysis to estimate the normal duration of human pregnancy. *Human Reproduction*, 16 (7), 1497-1500.

Smith, R., 1999. The timing of birth. *Scientific American*, 280 (3), 68-75.

Sorum, P.C., Stewart, T.R., Mullet, E., Gonzalez-Vallejo, C., Shim, J., Chasseigne, G., Sastre, M.T.M., Grenier, B. 2002. Does choosing a treatment depend on making a diagnosis? US and French physicians' decision-making about acute otitis media. *Medical Decision-making,* 22(5), 394 - 402.

Spiby, H., Green, J.M., Hucknall, C., Foster, H.R., Andrews, A., 2006a. *Labouring to better effect: studies of services for women in early labour*. Final report to the NHS Service Delivery and Organisation of Care Programme. York: Mother and Infant Research Unit.

Spiby, H., Crawshaw, S., Fyle, J., 2006b. ELSA Trial update: Support trial almost complete. *Midwives*, 9 (5), 184.

Stanovich, K.E., West, R., 2002. Individual differences in reasoning: Implications for the rationality debate. In: Gilovich, T., Griffin, D., Kahneman, D., eds. *Heuristics and Biases; the psychology of Intuitive Judgement.* Cambridge: Cambridge University Press. 421-440

Stewart, H.L., 1952. Duration of pregnancy and postmaturity. *Journal of the American Medical Association*, 148 (13), 1079 – 1083.

Stolte, J.F., 1994. The context of satisficing in vignette research. *Journal of Social Psychology,* 134, 727-733.

Streiner, D.L., Norman, G.R., 2003. *Health Measurement Scales: a practical guide to their development and use.* 3rd ed. Oxford: Oxford University Press.

Swets, J.A., 2000. Enhancing Diagnostic Decisions. In: Connolly, T., Arkes, H.R., Hammond, K.R. eds. *Judgement and Decision-making: An interdisciplinary Reader*. 2nd ed. Cambridge: Cambridge University Press. p66 - 82.

Tansey, E., Christie, D., 2000. *Looking at the unborn: historical aspects of obstetric ultrasound.* Wellcome Witness to Twentieth Century Medicine (5). Wellcome Trust, London, UK. Available from: http://ucl.ac.uk/histmed/PDFS/Publications/Witness/wit5.pdf

Tavakoli, M., Davies, H.T.O., Thomson, R., 2000. Decision analysis in evidence –based decision-making. *Journal of Evaluation in Clinical Practice*, 6 (2), 111-120.

Tew, M., 1990. *Safer Childbirth: A critical history of maternity care.* London: Chapman and Hall.

Thacker, S.B., Stroup, D., Chang, M., 2006. Continuous electronic heart rate monitoring for fetal assessment during labor. *Cochrane Database of Systematic Reviews,* Issue 3. Art. No.: CD000063. DOI: 10.1002/14651858.CD000063.pub2. (first published 1997*)*

Thomas, J., Paranjothy, S., 2001. *Royal College of Obstetricians and Gynaecologists Clinical Effectiveness Support Unit. National Sentinel Caesarean Section Audit Report.* London: RCOG Press.

Thompson, C., Cullum, N., McCaughan, D., Sheldon, T., Raynor, P., 2004. Nurses, information use, and clinical decision-making – the real world potential for evidence –based decisions in nursing. *Evidence Based Nursing*, 7 (3), 68-72.

Thompson, C., Dowding, D., 2001. Responding to uncertainty in nursing practice. *International Journal of Nursing Studies*, 38, 609-615.

Thornton, J.G., Lilford, R.J., 1994. Active management of labour: current knowledge and research issues. *British Medical Journal*, 309, 366 – 369.

Torgerson, D.J., 2001. Contamination in trials: is cluster randomisation the answer? *British Medical Journal*. 322, 355-357.

Treasure, T., MacRae, K.D., 1998. Minimisation: the platinum standard for trials. *British Medical Journal*, 317. 362-363.

Turnbull, D., Holmes, A., Shields, N., Cheyne, H., Twaddle, S., Gilmore, H. W., McGinley, M., Reid, M., Johnstone, I,. Greer, I., McIlwaine, G., Lunan, C.B.,1996. Randomised controlled trial of midwife managed care. *Lancet*, 348 (9022), 213-218.

Tracy, S.K., Tracy, M.B., 2003. Costing the cascade: estimating the cost of increased intervention in childbirth using population data. *British Journal of Obstetrics and Gynaecology*, 110, 717-724.

Tracy, S.K., Sullivan, E., Wang, Y.A., Black, D., Tracy, M., 2007. Birth outcomes associated with interventions in labour amongst low risk women: A population-based study. *Women and Birth*, 20, 41-48.

Trinder, L., 2000. A critical appraisal of Evidence-Based Practice. In: Trinder, L., Reynolds, S. eds. *Evidence –Based Practice: A critical appraisal.* Oxford: Blackwell Science. 212-241.

Turner, M.J., Webb, J.B., Gordon, H., 1986. Active management of labour in primigravidae. *Journal of Obstetrics and Gynaecology,* 7, 79-83.

Tversky. A., Kahneman, D., 1971. Belief in the law of small numbers. *Psychology Bulletin*, 76. 105-110.

Tversky, A., Kahneman, D., 1974. Judgement under uncertainty: Heuristics and biases. *Science*, 185 (4157), 1124-1131.

Tversky, A., Kahneman, D., 1983. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgement. *Psychology Review*, 90 (4), 293-315.

Tversky, A., Kahneman, D., 2000. Judgement under Uncertainty: heuristics and biases. In: Connolly, T., Arkes, H.R., Hammond, K.R., eds. *Judgement and Decision-making; an interdisciplinary reader.* 2nd ed. Cambridge: Cambridge University Press. p35-52.

Uldbjerg, N., Ulmsten, U., Ekman, G., 1983. Ripening of the human uterine cervix in terms of connective tissue biochemistry. *Clinical Obstetrics and Gynecology*, 26, 14-26.

Ulmsten, U., 1997. Onset and forces of term labour. *Acta obstetricia et gynecologica Scandinavica*, 76, 499- 514.

US Department of Health and Human Services, 1990. *Healthy People 2000. National health promotion and disease prevention objectives.* Washington: Public Health Service.

US Department of Health and Human Services, 1999. *National Vital Statistics Reports, Trends in the Attendant, Place, and Timing of Births, and in the Use of Obstetric Interventions: United States, 1989-17.* Centres for Disease Control and Prevention, National Center for Health Statistics. Volume 47 Number 27.

US Department of Health and Human Services, 2000. *Healthy People 2010. 2nd ed. With Understanding and Improving Health and Objectives for Improving Health.* Two vols. Washington: U.S. Government Printing Office. Available from: http://www.health.gov/healthypeople.

US Department of Health and Human Services, 2005. *National Vital Statistics Reports, Trends in Cesarean Rates for First Births and Repeat Cesarean Rates for Low-Risk Women: United States, 1990-2003.* Centres for Disease Control and Prevention, National Center for Health Statistics. Volume 54 Number 4.

van der Hulst, L.A.M., van Teijlingen, E.R., Bonsel, G.J., Eskes, M., Bleker, O.P., 2004. Does a pregnant woman's intended place of birth influence her attitude toward and occurrence of obstetric intervention? *Birth*, 31, 28- 33.

van Teijlingen. E., Hundley. V., 2005. Pilot studies in family planning and reproductive health care. *The Journal of Family Planning and Reproductive Health Care*, 31 (3), 219-221

Villar, J., Carroli, G., Zavaleta, N., Donner, A., Wojdyla, D., Faundes, A., Velazco, A., Bataglia, V., Langer, A., Naravaez, A., Valladares, E., Shah, A., Campodonico, L., Romero, M., Reynoso, S., de Padua, K.S., Giordano, D., Kublickas, M., Acosta. A., World Health Organisation 2005 Global Survey on Maternal and Perinatal Health Research Group., 2007. Maternal and neonatal individual risks and benefits associated with caesarean delivery: a multicentre prospective study. *British Medical Journal*, 335 (7628), 1025. Epub 2007 Oct 30.

Wagner, M., 1994. *Pursuing the birth machine: The search for appropriate birth technology.* Australia: ACE Graphics.

Wagner, M., 1998. Birth as a normal process. *Midwifery Today,* 47, 16-17.

Walker, A.E., Campbell, M.K., Grimshaw, J.M., and the TEMPEST group, 2000. A recruitment strategy for cluster randomised trials in secondary care settings. *Journal of Evaluation in Clinical Practice*, 6 (2), 185-192.

Walsh, D., El-Nemer, A., Downe, S. (2004) Risk, safety and the study of physiological birth. In: Downe, S., ed. *Normal childbirth: evidence and debate*. China: Churchill Livingstone. P103-120.

Wardle, S.A., Wright, P.J., Court, B.V., 1997. Knowledge of and preference for the DOMINO delivery option. *Midwifery*, 13 (3), 149-153.

Webb, C., Kevern, J., 2001. Focus groups as a research method: a critique of some aspects of their use in nursing research. *Journal of Advanced Nursing,* 33 (6), 798-805.

Wheble, A.K., Gilmer, M.D., Spencer, J.A., Sykes, G.S., 1989. Changes in fetal monitoring practice in the UK: 1977 – 1984. *British Journal of Obstetrics and Gynaecology*, 96 (10), 1140- 1147.

Wickstrom, G., Bendix, T., 2000. The "Hawthorne effect" – what did the original Hawthorne studies actually show? *Scandinavian Journal of Work, Environment and Health,* 26 (4), 363-367.

Williams, G.W., 2006. The other side of clinical trial monitoring; assuring data quality and procedural adherence. *Clinical Trials* 3 (6). 530-537.

Wood, L., Egger, M., Gluud, L.L., Schulz, K.F., Juni, P., Altman, D.G., Gluud, C., Martin, R.M., Wood, A.J.G., Sterne, J.A.C., 2008. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *British Medical Journal,* doi:10.1136/bmj.39465.451748.AD

World Health Organisation, 1985. Appropriate technology for birth. *Lancet*, 24 (2), (8452) 436-437.

World Health Organisation, 1996. *Care in normal birth: a practical guide*. In: Safe Motherhood. World Health Organisation, Geneva. Ch 1.

World Health Organisation, 1997. *Care in normal birth: A practical guide. Report of a technical working group.* World Health Organisation. Available from: http://www.who.int/reproductive-health/publications/MSM_96_24/MSM_96_24_table_of_contents.en.html

World Health Organisation, 2005. *The World Health Report 2005. Make every mother and child count*. Available from: http://www.who.int/whr/2005/whr2005_en.pdf

World Health Organisation, 2007. *International Classification of Diseases (ICD-10).* World Health Organisation. Available from: http://www.who.int/classifications/apps/icd/icd10online/

Wrigley, A.J., 1958. Postmaturity. *The Lancet*, 31st May, 1167 – 1168.

Zocco, J., Williams, M.J., Longobucco, D.B., Bernstein, B., 2007. A systems analysis of obstetric triage. *The Journal of Perinatal and Neonatal Nursing*, 21(4), 315-22.