



uOttawa

L'Université canadienne  
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES  
ET POSTDOCTORALES



uOttawa

L'Université canadienne  
Canada's university

FACULTY OF GRADUATE AND  
POSTDOCTORAL STUDIES

Sonya Kashyap

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Epidemiology)

GRADE / DEGRÉ

Department of Epidemiology and Community Medicine

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Assisted Reproductive Medicine: Systematic Reviews and Randomized Controlled Trials

TITRE DE LA THÈSE / TITLE OF THESIS

George Wells

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Douglas Coyle

Marc Rodger

Gary W. Slater

LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /  
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

**Assisted Reproductive Medicine:  
Systematic Reviews and Randomized Controlled Trials**

**By:  
Sonya Kashyap MD, FRCS( C )**

**Thesis submitted to the School of Graduate Studies and Research in the partial  
fulfillment of the requirements for the MSc degree in Epidemiology**

**University of Ottawa**

**Thesis Supervisor:  
Dr. George A. Wells**



Library and  
Archives Canada

Bibliothèque et  
Archives Canada

Published Heritage  
Branch

Direction du  
Patrimoine de l'édition

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*

*ISBN: 0-494-11310-3*

*Our file* *Notre référence*

*ISBN: 0-494-11310-3*

**NOTICE:**

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

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

**AVIS:**

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**



## **Abstract**

The objective for this thesis was to propose a question in the area of reproductive medicine that could be addressed by a clinical trial. In preparation for such a trial we conducted a systematic review of the topic.

Our first question was whether metformin is of benefit compared to clomiphene citrate for ovulation induction and achievement of pregnancy in women with polycystic ovarian syndrome (PCOS). We conducted a systematic review (SR) and meta-analysis of the subject before designing and implementing a randomized controlled trial (RCT). The RCT was terminated for recruitment issues. Our PCOS SR led us to develop a novel ovarian stimulation protocol for poor responders. We used the lessons learned from our failed RCT and another systematic review to design and conduct a feasibility randomized controlled trial on the use of aromatase inhibitors to improve pregnancy outcomes for in-vitro fertilization in poor responders. We successfully completed the pilot study and found a trend towards increased clinical pregnancy rates for patients who received the study versus the standard protocol. The pilot study provides recommendations for the definitive trial.

## **Acknowledgements:**

First and foremost, I am greatly indebted to my thesis supervisor, Dr. George A. Wells, for his invaluable mentorship, teaching, unfaltering support and guidance. His commitment to this project was limitless, both in effort and time, and his encouragement motivated me through the most discouraging periods. His technical wizardry translated the most complicated concepts into simplicities. He is truly an outstanding mentor.

Next, I would like to thank my clinical mentor and fellowship program director, Dr. Zev Rosenwaks from Center for Reproductive Medicine at Cornell University Medical College in New York, New York. He provided not only the environment, commitment and resources necessary to complete the clinical aspects of this project but also invaluable clinical insight and moral support without which this project would not have been possible. He is the definition of a “great doctor”. He encouraged me to think outside the box and to be relentless in the pursuit of excellence for my patients.

I would also like to thank Dr. Owen Davis for his clinical guidance, mentorship, and critique of the protocol prior to submission to the Institutional Review Board.

Thanks is also due to Mary-Anne Williams Pittman, the research nurse without whom subject identification, randomization and consent processes, as well as data collection would have been infinitely more onerous. Her commitment to this project is greatly appreciated.

The cooperation of the entire medical staff at the Center for Reproductive Medicine including attending physicians, fellows, embryologists, andrologists, nurses and administrative staff was instrumental in the completion of this thesis. I would like to extend my appreciation to the other attending physicians who allowed their patients to participate: Drs. Pak Chung, Glenn Schattman, Dan Goldschlag, Steven Spandorfer, Isaac Kligman, and Ina Cholst.

Finally, I would like to thank my family for their unconditional faith, encouragement, and patience through my many years of postgraduate education. Thank you to my mother whose example of strength, tenacity, honesty and kindness in the face of adversity taught me that the human spirit can overcome any obstacle.



## Table of contents

	Page
Abstract.....	i
Acknowledgements.....	ii
Table of Contents.....	iv
List of Tables.....	viii
List of Figures and Illustrations.....	ix
List of Abbreviation and Acronyms.....	x

### Chapter 1: Thesis Overview

1.1	Introduction.....	1
1.2	Thesis Objectives.....	2
1.3	Thesis Progression.....	3

### Chapter 2: PCOS Study

**Systematic review and randomized controlled trial of metformin versus clomiphene citrate for ovulation induction and achievement of pregnancy in women with polycystic ovarian syndrome**

2.1	Introduction.....	7
2.1.1	Background	
2.1.1.1	Epidemiology.....	8
2.1.1.2	Pathophysiology.....	10
2.1.1.3	Clinical Treatment Options, Success Rates, and Evidence.....	12
2.1.1.4	Rationale for Proposed Study.....	16
2.1.2	Objectives.....	17
2.2	PCOS Systematic Review of Metformin vs CC	
2.2.1	Methods for Conducting a Systematic Review.....	18
2.2.1.1	Search Strategy.....	18
2.2.1.2	Inclusion/ Exclusion.....	21
2.2.1.3	Quality Assessment.....	21

2.2.1.4	Study Inclusion (PICOS).....	21
2.2.1.5	Quantitative Methods.....	22
2.2.2	Results of PCOS Systematic Review	
2.2.2.1	Study Inclusion.....	22
2.2.2.2	Characteristics of Included studies.....	23
2.2.2.3	Meta-analysis.....	29
2.2.3	Discussion of PCOS Systematic Review.....	38
2.3	Randomized controlled trial of Metformin vs CC in PCOS	
2.3.1	Methods for Developing a Randomized Controlled Trial.....	43
2.3.1.2	PCOS RCT Protocol.....	44
2.3.2	Results of PCOS RCT	
2.3.2.1	Study Implementation.....	55
2.3.2.2	Study Termination.....	56
2.3.3	Discussion of PCOS RCT .....	56

### **Chapter 3: Aromatase Inhibitors in Poor Responders and IVF**

**Systematic review and randomized controlled trial feasibility study for novel use of aromatase inhibitors to improve pregnancy outcomes in in-vitro fertilization of poor responders**

3.1	Introduction.....	60
3.1.1	Background.....	62
3.1.1.1	Pathophysiology of Poor Responders.....	64
3.1.1.2	Epidemiology of Poor Responders.....	65
3.1.1.3	Clinical Treatment Options, Success Rates & Evidence.....	66
3.1.1.4	Rationale for a Feasibility Study.....	68
3.1.2	Objectives.....	70
3.2	Systematic Review for Aromatase Inhibitors for Poor Responders	
3.2.1	Methods for Conducting a Systematic Review	
3.2.1.1	Search Strategy.....	71

3.2.1.2	<b>Inclusion/ Exclusion.....</b>	<b>72</b>
3.2.1.3	<b>Quality Assessment.....</b>	<b>73</b>
3.2.1.4	<b>Study Inclusion (PICOS).....</b>	<b>73</b>
3.2.1.5	<b>Quantitative Methods.....</b>	<b>74</b>
3.2.2	<b>Results of Systematic review</b>	
3.2.2.1	<b>Study Inclusion.....</b>	<b>74</b>
3.2.2.2	<b>Characteristics of Included studies.....</b>	<b>75</b>
3.2.2.3	<b>Meta-analysis.....</b>	<b>75</b>
3.2.3	<b>Discussion of the Systematic Review.....</b>	<b>75</b>
3.3	<b>Poor Responder RCT</b>	
3.3.1	<b>RCT Protocol for Aromatase Inhibitors in Poor Responders.....</b>	<b>76</b>
3.3.2	<b>Results of the Poor Responder Feasibility Study.....</b>	<b>88</b>
3.3.3.1	<b>Study Outcomes.....</b>	<b>88</b>
3.5.3.2	<b>Feasibility Outcomes.....</b>	<b>100</b>
3.3.3	<b>Discussion</b>	
3.3.3.1	<b>Discussion of Results of the Pilot Study.....</b>	<b>103</b>
3.3.3.2	<b>Assessment of Feasibility Study Objectives.....</b>	<b>105</b>
3.3.3.2.1	<b>Recruitment.....</b>	<b>105</b>
3.3.3.2.2	<b>Randomization.....</b>	<b>109</b>
3.3.3.2.3	<b>Determination of the Primary Outcome &amp; Sample Sample Size</b>	<b>110</b>
3.3.3.2.4	<b>Logistics of Study Implementation.....</b>	<b>112</b>
3.3.3.3	<b>Ethics.....</b>	<b>115</b>
3.3.3.4	<b>Summary.....</b>	<b>116</b>

## **Chapter 4: Discussion of the Thesis**

4.1	<b>Evidence Based Medicine.....</b>	<b>117</b>
4.2	<b>Challenges of Developing and Implementing RCTs in</b>	

	<b>Reproductive Medicine.....</b>	<b>118</b>
<b>4.3</b>	<b>The Challenge of MCID.....</b>	<b>119</b>
<b>4.4</b>	<b>Recommendations.....</b>	<b>121</b>

**Appendices**

<b>Appendix A</b>	<b>Jadad Quality Scale</b>
<b>Appendix B1</b>	<b>PCOS RCT Flowchart</b>
<b>Appendix B2</b>	<b>PCOS RCT Consent Form</b>
<b>Appendix C1</b>	<b>Published PCOS Meta-Analysis – Insulin Sensitizing Agents as Primary Therapy for Polycystic Ovarian Syndrome</b>
<b>Appendix C2</b>	<b>Oxford Center for Evidence Based Medicine- Levels of Evidence</b>
<b>Appendix D</b>	<b>Diagram for RCT Topic Change</b>
<b>Appendix E</b>	<b>Poor Responder RCT Consent Form</b>

Table 1	A step by step approach to ovulation induction in women with PCOS	12
---------	-------------------------------------------------------------------	----

Table 2	PCOS SR search strategy	<b>25</b> <b>Page</b>
---------	-------------------------	--------------------------

Table 3	Quality assessment by Jadad scale of PCOS RCTs	27
Table 4	Characteristics of PCOS RCTs	28
Table 5	Comparison of 2 meta-analyses for PCOS	72
Table 6	Poor Responder search strategy	89
Table 7	Poor Responder patient demographics	92
Table 8	Frequency of each inclusion criteria in overall, standard and letrozole groups.	93
Table 9	Frequency of participants who met greater than 1 inclusion criteria.	94
Table 10	Comparison of patients who were randomized but did not start stimulation versus patients who did start stimulation.	95
Table 11	Clinical pregnancy rates	93
Table 12	Secondary outcomes in subjects who began treatment	96
Table 13	Characteristics of subjects who achieved pregnancy	102

### **List of Tables**

## List of Figures

		<b>Page</b>
Figure 1	Progression of thesis	4
Figure 2	Study inclusion	25
Figure 3	Metformin vs placebo in infertile PCOS patients	32
Figure 4	Metformin vs placebo for cycle regulation in PCOS patients	33
Figure 5	Metformin plus clomiphene citrate vs placebo plus clomiphene citrate	34
Figure 6	Funnel plot of metformin vs placebo in PCOS patients	35
Figure 7	Funnel plot of metformin vs placebo in PCOS patients (II)	36
Figure 8	Funnel plot of metformin +CC vs placebo + CC in PCOS patients	37
Figure 9	Poor Responder ovarian stimulation protocols for IVF	81
Figure 10	Poor responder RCT protocol	94
Figure 11	Flowchart of the randomized controlled trial (Poor Responder)	90
Figure 12	Frequency of number of inclusion criteria met	91

## List of Abbreviations and Acronyms

ART	Assisted reproductive technology
BMI	Body mass index
CC	Clomiphene citrate
CDC	Centers for Disease Control
CI	Confidence Interval
COH-IUI	Controlled ovarian hyperstimulation and in-utero insemination
CRMI	Center for Reproductive Medicine and Infertility at Cornell University
D2	Day 2
E2	Estradiol
ET	Embryos transferred
FDA	Food and Drug Administration
FSH	Follicular Stimulating Hormone
GnRH	Gonadotropin Releasing Hormone
GTPAL	Gravidity, term, preterm, aborta, live (Standard annotations for pregnancy history)
hCG	Human Chorionic Gonadotropin
hMG	Human menopausal gonadotropin
HMO	Health Management Organizations
ICSI	Intracytoplasmic sperm injection
IGF-1	Insulin-like growth factor
IR	Implantation rate
IVF	In-vitro fertilization
LH	Luteinizing Hormone
MCID	Minimal clinically important difference
NICHD	National Institutes for Child Health and Human Development
NICU	Neonatal intensive care unit
NIH	National Insitutes of Health
NOS	Newcastle scale for Observational Studies
OHSS	Ovarian hyperstimulation Syndrome
PCOS	Polycystic ovarian syndrome
PCOD	Polycystic ovarian disease
PI	Principal investigator
RCT	Randomized controlled trial
RR	Relative risk
SART	Society for Assisted Reproductive Technology
SHBG	Sex hormone binding globulin
SR	Systematic review
TESE	Testicular epididymal sperm extraction
US	United States

# Chapter 1 Thesis Overview

## 1.1 Introduction

Infertility, commonly defined as the inability to conceive after one year of unprotected intercourse, affects 10- 15% of couples. Although the incidence of infertility has remained stable, the use of infertility drugs has increased rapidly. It is estimated that 2.5% of all North American births are the result of artificial reproductive technology. (1)

The Food and Drug Administration (FDA) registered clomiphene citrate in 1967. Gonadotropins were registered in 1969. Prior to that time, the hormonal treatment of infertility involved the use of estrogens, progestins, oral contraceptive medications, and pituitary radiation. Since the first human birth from an in-vitro fertilized conceptus occurred in 1978, utilization of in-vitro fertilization (IVF) has expanded rapidly. In 2001, 40687 live babies were born from assisted reproductive technology (ART) cycles in the United States (US) alone.( <http://www.cdc.gov/reproductivehealth/ART01/acknowl.htm>)

Collins estimated that the cost of a cycle of in-vitro fertilization, without assisted micromanipulation, averaged almost \$10000 in the United States in the year 2001.(2) Fertility treatment success rates vary according to patient age, infertility diagnosis, clinic expertise, and clinic size but assisted reproductive technology and infertility is an area lacking in evidence based medicine. Several challenges present. Therapy is often invasive, expensive, and time consuming. Since most therapies require patients to pay out of pocket in most districts in North America, patients and physicians are often unwilling to participate in randomized controlled trials. Therapies that may be of benefit, particularly in patients with poor prognoses, are often implemented in the absence of alternatives. Most studies report surrogate outcomes such as oocyte and embryo yields or biochemical pregnancy rates rather than clinically relevant pregnancies or delivery rates. This may be due differing perspectives of physicians who treat infertility versus those who care for these patients during pregnancy and delivery.



The cost for a cycle of controlled ovarian hyperstimulation – in utero insemination (COH-IUI) is more difficult to derive but it is still significant. The success rates for COH-IUI range from 5 –33%. ([http://www.nichd.nih.gov/cpr/rs/rep\\_medicine.htm](http://www.nichd.nih.gov/cpr/rs/rep_medicine.htm))

Until now, basic science has contributed most of the research around ART. These advances increased the efficiency of ART. Despite rapidly increasing utilization of reproductive technologies, very little rigorous evaluation of different treatment modalities is performed before many of these treatments become standard of care. A Medline search reported 449 papers with “randomized controlled trials” as a MeSH heading, including editorials and letters. Many of the reported randomized controlled trials (RCTs) were not properly designed or conducted.

As of 2001, only three states in the United States mandated complete IVF coverage.(3) Less invasive treatments (ie COH-IUI) have varying levels of coverage in both Canada and the US. Nonetheless, the direct costs of treatment to the patients are high. Much debate has focused on the financial and emotional burdens of couples suffering from infertility when and how, after repeated failures, they should be counseled to discontinue treatment.(4, 5) More responsibility should also be placed on the reproductive medicine community to prospectively evaluate treatments in a systematic way, for given patient groups, to determine optimal treatment and likely success rates at initiation of therapy.

Therefore, our major goals in this thesis were: 1) to propose a treatment question in an area of reproductive medicine that could be evaluated by a randomized controlled trial and to conduct a systematic review of this topic which would then aid in the design of such a trial; and 2) to design and implement the trial with the expectation of answering the question and gaining potential knowledge on the difficulties of developing and conducting such studies.

## **1.2 Objectives**

The overall goals of this thesis from a methodologic viewpoint were as follows: to propose a concise and explicit question in the area of reproductive medicine which may be

answered by a randomized controlled trial; 2) to perform a systematic review on the above topic, using QUoRUM and MOOSE guidelines, in preparation for designing and conducting the randomized controlled trial; 3) to design a randomized controlled trial using robust methods following CIHR guidelines; and 4) to implement and conduct the randomized controlled trial/ feasibility study.(6-8)

The specific objectives were to answer two questions:

“Is metformin beneficial compared to clomiphene citrate for achievement of pregnancy in patients with PCOS?” and

“Are aromatase inhibitors, in combination with gonadotropins, beneficial compared to standard, aggressive protocols for IVF treatment of poor responders?.

We proceeded to do this by designing and implementing systematic reviews on each topic, followed by designing and implementing randomized controlled trials/feasibility studies which were based on the results of the systematic reviews. Our study progressed from the first topic to the second topic as a result of issues raised during the first systematic review followed by recruitment issues in the first RCT.

### **1.3 Thesis Progression**

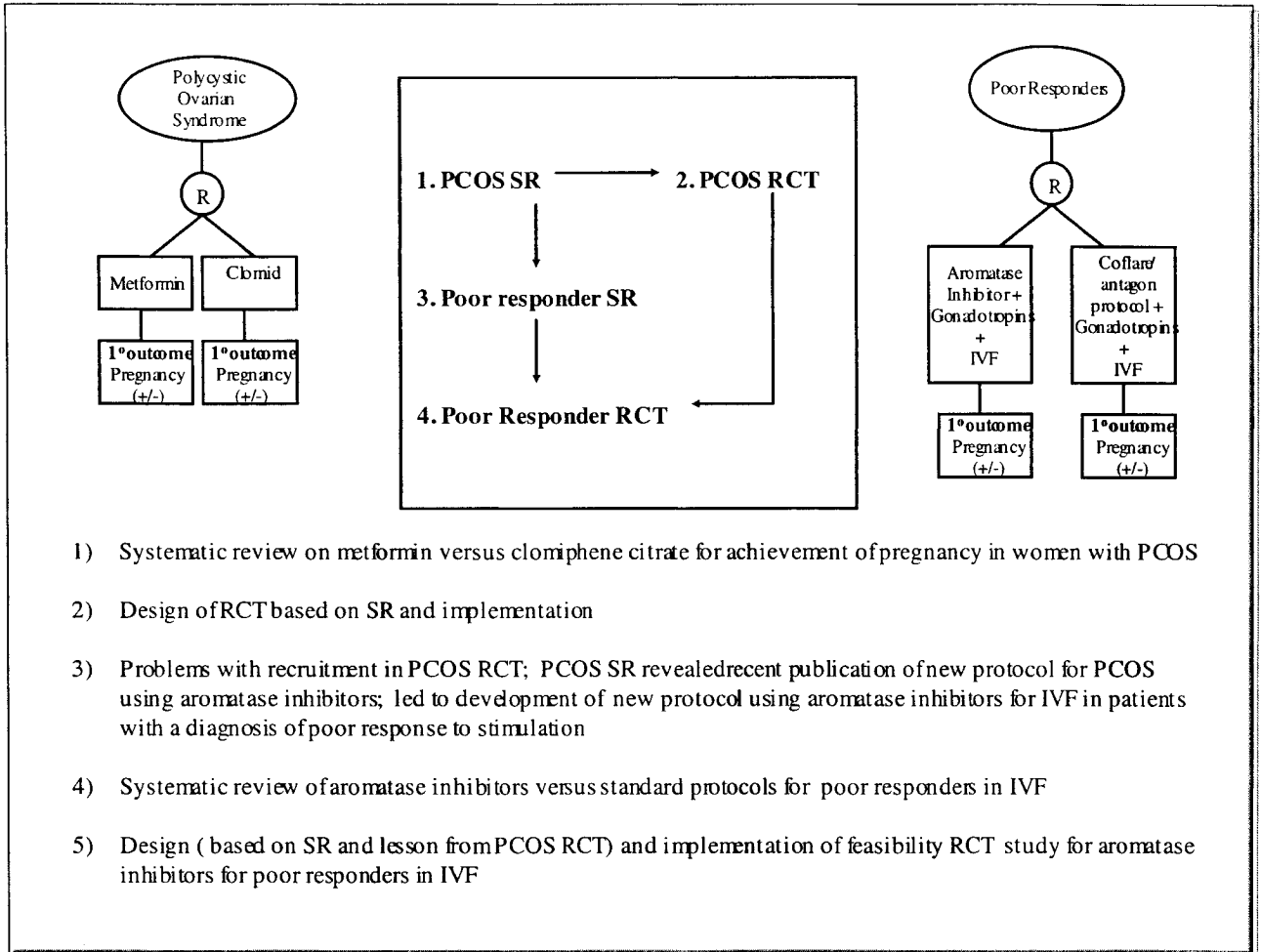
The overall goal of this thesis was to perform a systematic review (SR) on a question in reproductive medicine and then design a randomized controlled trial study based on the results of the systematic review. An overview of the progression of the thesis is given in Figure 1.

We began this thesis with the question “is metformin as effective a treatment as clomiphene citrate (CC) in the first line treatment of infertility in women with polycystic ovarian syndrome (PCOS) when the primary outcome considered is pregnancy?”

Polycystic ovarian syndrome is the most common endocrine disorder in women. PCOS affects approximately 5-10% of women and up to 75% of anovulatory women.

Anovulation is responsible for approximately 40% of cases of infertility. (9, 10)

**Figure 1: Progression of thesis**



Currently, standard treatment for induction of ovulation in these patients is clomiphene citrate (CC). Since the first observational, cohort studies demonstrated resumption of menstrual cycles in 21 of 22 patients with PCOS, many centres employ metformin as primary therapy.(11) Several studies have been conducted on small samples of women

with PCOS with the primary outcomes ranging from lowering of total insulin and androgen levels to reinstatement of ovulation. Most of these studies were observational and many retrospective. The RCTs that do exist compare metformin to placebo and use ovulation rather than pregnancy as a primary outcome. There are no direct, head to head RCTs of metformin versus CC in women with PCOS.

Chapter 2 outlines the systematic review, using Cochrane, MOOSE and QUORUM guidelines, for metformin versus clomiphene citrate for ovulation induction and achievement of pregnancy in women with PCOS.(6, 7) We then employed CIHR and CONSORT guidelines to develop and implement a randomized controlled trial on this topic.(12) A detailed protocol is found in the results section of chapter 2. The results of both the systematic review and RCT are presented at the end of chapter 2.

During the process of conducting the systematic review on medical induction of ovulation for patients with polycystic ovarian syndrome, a novel paper was published suggesting that aromatase inhibitors may benefit this population.(13) However, our review of the physiology of aromatase inhibitors suggested to us that this category of drug may be more beneficial in another special group of patients called poor responders. Poor responders are those patients who have a low probability to achieve pregnancy even with in-vitro fertilization. Special characteristics of poor responders may include one or more of the following: age over 40 years; estradiol less than 1000 pg/ml with previous IVF stimulation; less than 5 follicles/oocytes obtained at retrieval; history of previous elevated estradiol (> 75 pg.ml) or day2 and /or 3 follicle stimulating hormone (FSH) (> 12 mIU/ml DPC assay/ > 20 mIU/ml RIA assay); and/or history of previous cancellation of IVF stimulation for poor response.

We therefore conducted a search for clinical trials which investigated the effect of aromatase inhibitors in poor responders both for in-utero insemination and for in-vitro fertilization. We found several studies, including one randomized controlled trial, which assessed aromatase inhibitors in PCOS patients. (14)We did not find any studies looking at poor responders and IVF except for one abstract of a case-series study which

investigated the use of aromatase inhibitors for the purpose of ovulation induction in women who were called “low responders”.(15)

We experienced difficulty with recruitment for the PCOS study and discovered this was due to a new decision at the study site (Cornell Medical College) to discontinue acceptance of three types of insurance. This decision significantly reduced not only the total number of ovulation induction cycles we had been seeing but also the number of PCOS patients who were presenting for ovulation induction. This resulted in discontinuation of the PCOS RCT.

In contrast, we have been steadily seeing an increase in IVF patients for which patients pay privately. Cornell has an international reputation for having among the highest success rates for any IVF clinic, particularly with poor responders. We therefore have referrals from all over the world. For this reason and the fact that we know we see at least 200 poor responders every three months, we decided to initiate a second randomized controlled trial in poor responders.

Based on the physiology of aromatase inhibitors, we felt that aromatase inhibitors would be appropriate for a group of IVF patients called “poor responders”. We therefore developed a new protocol for these patients called “poor responders” and performed a formal systematic review of this topic. The subsequent protocol was novel, unpublished and attractive to many of the clinicians. We therefore designed and implemented a new RCT of aromatase inhibitors in IVF for poor responders, based on the systematic review and lessons learned from the PCOS RCT. The methods and results for the systematic review, protocol development and RCT implementation as a feasibility study, are presented in chapter 3.

## Chapter 2: PCOS Study

### 2.1 Introduction

Polycystic ovarian syndrome (PCOS) was first described by Stein and Levanthal in 1935 as a classic triad of amenorrhea, obesity, and hirsutism. Since then, much has been learned about the pathophysiology of the syndrome and these advances have led to increased treatment options. Perhaps the most significant advance has been the link to hyperinsulinemia; the interaction of hyperinsulinemia with hyperandrogenism leading to chronic anovulation. Traditionally, first line treatment for PCOS has involved ovulation induction with clomiphene citrate (CC).(16) In 1994, Velasquez published the first paper to suggest the use of metformin for resumption of menstrual cyclicality in women with PCOS. In a prospective uncontrolled cohort of 22 PCOS women, 21 (95.7%) resumed menstrual cyclicality and 13 of 15 women tested had ovulatory progesterone levels (3.1–28 ng/ml).(11)

The debate about whether or not metformin should be first line therapy in PCOS women, who wish to conceive, is ongoing. There are many proponents for the use of metformin, however, most agree that the evidence to date arises mainly from poorly designed observational studies and that the few randomized controlled trials (RCTs) that exist were not appropriately designed regarding, sample size, randomization methods, and primary outcome.(17-19) In particular, no trials exist that assess pregnancy as a primary outcome.

CC is the traditional, first line medical therapy for patients with PCOS who wish to conceive. CC has remained standard therapy in this group of patients for over three decades. Ovulation rates are reported to be as high as 70% in some papers with pregnancy rates approaching 30-40%. However, CC induces superovulation (> 1 mature follicle) thereby placing patients at moderate risk for multiple birth (twins in 10% of cases and triplets in < 1%) and theoretical, although undocumented, risks of ovarian hyperstimulation. CC acts as an antiestrogen at the level of the endometrium and cervical mucus which may have adversely affect attempts at conception. While most physicians

administer CC for use in combination with natural intercourse for simplicity and compliance, National Institutes of Health (NIH) data supports improved results with in-utero insemination. Also, prolonged use of CC (> 12 months) has been associated with an increased risk of ovarian cancer in infertile treated women versus infertile untreated women.(20, 21) Potential advantages of metformin over CC include monofollicular ovulation induction, cycle regulation absence of anti-steroidogenic action, avoidance of ovarian cancer risk and the need for in-utero insemination.

A meta-analysis of CC versus placebo has been previously published.(22) Compared with placebo, CC was associated with an OR of 6.82 (95% CI 3.92, 11.85) for ovulation and OR 3.41 (4.23, 9.48) for pregnancy. However, the analyses were performed on studies of cross over design rather than parallel group RCT. Cross over trials are not the appropriate study designs in this group.

However, there are no head to head RCTs of (CC) versus metformin in women with PCOS for either metabolic, ovulation, or pregnancy outcomes.

This chapter presents the design and implementation of a randomized controlled trial to answer the question “is metformin of benefit to clomiphene citrate in the primary treatment of women with PCOS who wish to conceive?”

## **2.1.1 Background**

### **2.1.1.1 Epidemiology**

Polycystic ovarian syndrome is the most common endocrine disorder in women. PCOS affects approximately 5-10% of reproductive aged women, as determined by using the NIH criteria in prospective cohorts of unselected populations, and up to 20% of infertile women.(23-26) Anovulation is responsible for approximately 40% of cases of infertility and some estimates suggest that 75% of anovulation may be attributed to PCOS.(27, 28) The 2001 CDC National Report and Fertility Clinics Summary reported that 6% of 77102 fresh, nondonor egg IVF cycles (approximately 4626 cases) were started for a singular diagnosis of ovulatory dysfunction. Other causes of ovulatory dysfunction may include anovulation secondary to hypothalamic-pituitary issues such as hyperprolactinemia,

hypo/hyperthyroidism, and stress or anorexia induced anovulation. PCOS, however, likely contributes the bulk of these cases and since most cases of PCOS may be successfully treated with ovulation induction by metformin, clomiphene citrate, or gonadotropins (plus/minus insemination), this number greatly underestimates the burden of disease secondary to infertility.

There are several definitions of PCOS: Stein-Levanthal, Adams criteria, WHO type II anovulation, NIH definitions and most recently the ESHRE-ASRM Rotterdam consensus. Stein and Levanthal first described the syndrome in 1939 as a classic triad of amenorrhea, obesity and hirsutism. They required histological confirmation by direct visualization and ovarian biopsy to confirm the diagnosis. Adams criteria is based mostly on ovarian morphology and sonographic appearance.(27) Although most experts agree that chronic hyperandrogenism and menstrual dysfunction are important hallmarks of the syndrome, specific criteria for the diagnosis of PCOS remain elusive particularly in relation to clinical trials.(29) The clinical manifestations of hyperandrogenism may not always accompany elevated androgen levels since genetic susceptibility to these hormones at the target organ is also required. Hyperinsulinemia has been documented to occur in 50-100% of obese PCOS patients and as high as 22% of lean PCOS patients.(30)

The most recent NIH consensus conference on PCOS in 1990 defined the syndrome as chronic anovulation with hyperandrogenism after exclusion of other causes of infertility. Ratios of luteinizing hormone (LH) to follicular stimulating hormone (FSH), ultrasound criteria and hyperinsulinemia were so far not included in this definition. The 2003 Rotterdam ESHRE/ASRM consensus on PCOS attempted to delineate a clearer definition of PCOS but also stated that “ PCOS remains a syndrome and no single diagnostic criterion (such as hyperandrogenism or polycystic ovaries) is sufficient for clinical diagnosis. PCOS also remains a diagnosis of exclusion.” (31) Indeed, even the prerequisite of oligomenorrhea has been challenged in the mildest form of the syndrome.(29, 32) Nevertheless, the consensus stated that the revised diagnostic criteria include two of the following three findings: 1) oligo or anovulation, 2) clinical and/or biochemical signs of hyperandrogenism, 3) polycystic ovaries and exclusion of other etiologies.(31) Clearly, as



the Rotterdam proceedings suggested, the definition of PCOS may be flexible but the criteria employed in a study should directly address the interventions and outcomes assessed. Although the revised Rotterdam criteria are more appropriate for our study, we used the NIH guidelines as the former were not published at the beginning of our study. Ovulation and pregnancy rates, not metabolic parameters, were to be assessed in our study.

Patients with polycystic ovarian syndrome have been shown to have genetic susceptibility for this condition. The co-incidence of hyperandrogenemia in sisters of affected patients is almost 50% and some authors have suggested that the inheritance of PCOS appear to be autosomal dominant.(33) Family studies have suggested an association and linkage of the insulin gene with PCOS.(34)

#### **2.1.1.2 Pathophysiology**

Patients with PCOS are at risk for significant health disturbances throughout their lives. These patients are at increased risk for early adult onset diabetes, cardiovascular disease, endometrial cancer and possibly breast cancer. However, the most common presenting complaint early in life is infertility. Several theories exist regarding the pathogenesis and pathophysiology of PCOS. Inherent hypothalamic-pituitary disorders affecting the gonadotropin releasing hormone (GnRH) pulse generator, intrinsic alterations of ovarian function possibly due to genetic abnormalities of CYP 17 and/or CYP 11 systems, and insulin resistance are the most popular theories. Women with PCOS are known to have a larger number of antral follicles or an excessive endowment of follicles from fetal life which contributes to later clinical manifestations.(35, 36) We will concentrate on the relationship between insulin resistance, hyperinsulinemia, the interaction between hyperinsulinemia and hyperandrogenism and how this relationship leads to chronic anovulation.

Patients with PCOS have been demonstrated to have increased GnRH pulse frequency and amplitude that result in elevated levels of LH. The adrenal gland, in PCOS patients, increases production of androstenedione through increased 17, 20-lyase activity.

Androstendione in turn is peripherally converted into estrone. Ovarian production of estradiol is not increased, however, the total perceived estrogen levels are increased because of : 1) increased estrone and 2) decreased sex-hormone binding globulin (SHBG). Elevation of total estrogen levels causes negative feedback on FSH thereby preventing recruitment of antral follicles to form a dominant follicle, required for ovulation. Both increased androgen and insulin both reduce hepatic production of sex hormone binding globulin (SHBG). Increased androgen levels in patients with PCOS have several sources: 1) the adrenal gland has demonstrated increased 17, 20-lyase activity which increases production of androstenedione; 2) the cycle of decreased SHBG, due both to direct androgen and insulin suppressive effects, also increases free testosterone levels; 3) increased LH levels stimulate ovarian thecal cells to further produce increased androgens which in turn prevent the follicles from developing an estrogen dominant environment required for ovulation; 4) failure to ovulate further increases the proportion of luteinized thecal cells available for androgen production; 5) insulin resistance results in increased insulin levels that have two primary effects which also contribute to the androgen pool : a) decreased hepatic production of SHBG and insulin-like growth factor binding protein (IGF-1); and b) direct stimulation of both IGF-1 and insulin receptors on ovarian thecal cells which produce androgens. High concentrations of androgens further inhibit estrogen dominance by inhibiting aromatase action required for conversion of androgens to estrogen in the granulosa cells.

Defects in insulin post-receptor binding and excessive serine phosphorylation have been shown to be responsible for the insulin resistance in many of these patients. Attempts to determine hyperinsulinemia in these patients have been fraught with difficulty. The gold standard euglycemic clamp is invasive, tedious, and rarely justified. Other tests such as the oral glucose tolerance test and fasting glucose:insulin ratio( $<4.5$ ) are specific but not sensitive. Nevertheless, the fasting glucose: insulin ratios in combination with clinical predictors of hyperinsulinemia are currently the standard of care in clinical assessment of these patients.

The evidence that hyperinsulinemia causes hyperandrogenism, which in turn affects fertility, is more convincing than arguments that hyperandrogenism causes hyperinsulinemia. Following insulin infusion, androgen levels rise.(37) Conversely, treatment of hyperinsulinemia with insulin sensitizing agents or weight loss (5% or BMI< 27) decreases androgen levels.(38, 39) Medically or surgically induced menopause decreases circulating androgen levels without any affect on serum insulin levels.(40)

### 2.1.1.3 Clinical treatment options, success rate and evidence

As knowledge about the pathophysiology of PCOS expands, so do treatment options. The overall prognosis to achieve pregnancy in these patients is very good. However, the order in which such treatment options should be pursued may vary according to individual patient history and therapeutic objectives. Kim and colleagues published a suggested algorithm for induction of ovulation and achievement of pregnancy in women with PCOS.(41) (Table 1)

**Table 1: A step by step approach to ovulation induction in women with PCOS**

Step	Approach
1	If BMI is elevated, loss of at least 5% of current body weight
2	Ovulation induction with clomiphene (+/- glucocorticoid if elevated DHEAS)
3	Insulin sensitizer as a single agent
4	Insulin sensitizer in combination with clomiphene
5	Gonadotropin therapy
6	Insulin sensitizer in combination with gonadotropin therapy
7	Ovarian surgery
8	IVF

Kim, Insulin sensitizers for infertility. Fertil Steril 2000.(41)

The first established treatment of ovarian wedge resection was pioneered after Stein and Levanthal described PCOS to be a triad of obesity, hirsutism and amenorrhea that could be macroscopically visualized as an ovarian morphologic disorder and histologically diagnosed.(42, 43) The principle behind wedge resection is to remove a section of ovarian

tissue in order to reduce excess androgen production and therefore revert anovulatory cycles to spontaneous ovulation. Laparoscopic ovarian drilling has increased the efficiency of this procedure by decreasing potential operative complications, recovery time, and adhesion formation.(44, 45) Advantages to these treatments include resumption of menstrual cyclicity, lower androgen levels and longer term therapeutic benefits which are not limited to fertility treatment. Nonetheless, surgical treatment as primary therapy has fallen out of favour.

Weight loss (minimum 5% current weight) has also been proven to be of benefit to restore ovulatory cycles in obese PCOS women. (46, 47) Pasquali et al. published an uncontrolled cohort of obese women with PCOS ( criteria: obesity, amenorrhea and clinical/ biochemical hyperandrogenism) who were treated with a hypocaloric diet for an average of eight months. Biochemical and clinical features significantly improved. Mean plasma testosterone levels and LH levels significantly decreased (  $p < 0.0001$ ). Hirsutism scores improved in greater than 50% of subjects and 8 of 20 women resumed menstrual cyclicity and 4 became pregnant. Weight loss averaged  $9.7 \pm 3.1$  kg overall and  $1.35 \pm 0.56$  kg/month.(48) A subsequent study evaluated a cohort of 67 obese, infertile women with PCOS who completed a six month regimen of diet and lifestyle alterations. The average weight loss was  $10.2 \text{ kg/m}^2$ , 60 of 67 anovulatory patients resumed spontaneous ovulation and 52 of 67 achieved pregnancy. Anovulatory women who attended more than 66% of the sessions resumed spontaneous ovulation.(49) Taken together, in addition to other positive studies, this data suggests that diet and lifestyle alterations are highly efficacious. Four studies which evaluated such treatment were reported as RCTs but in fact were not true RCTS in that the allocation of treatment was often sequential or at physician discretion rather than being truly randomized. Indeed, several authors have suggested that lifestyle modification should be the primary therapy in this patient group.(41, 49) Nevertheless, such programs are difficult to implement and even more difficult to maintain. Such treatments may not be of benefit in lean PCOS patients. Therefore other treatment modalities are necessary.

Clomiphene citrate (CC) was FDA approved for the treatment of infertility in 1967. CC accounts for two thirds of fertility drugs prescribed in the United States.(50) At that time, numerous RCTs were conducted to evaluate the efficacy of CC in the treatment of infertile, oligomenorrheic women. Indeed, CC became the recommended initial therapy for such patients. Currently, standard treatment for ovulation induction in PCOS patients is CC. The success rate of CC varies; 80% percent ovulation and 30-40% pregnancy rates have been reported.(51-55) Seventy five percent of pregnancies are achieved within the first three months of an ovulatory dose of CC.(52) The proposed mechanism of action is that CC acts at the hypothalamus as a “weak estrogen” and therefore the pituitary perceives a lower estradiol level and increases FSH to recruit antral follicles. However, CC also has an anti-estrogenic effect on the endometrial lining and cervical mucous. The incidence of multiple pregnancy (mostly twins) with CC is 4 - 10%.(54, 56) There is also a 1% risk of ovarian hyperstimulation (OHSS) after gonadotropin therapy and although this may be much less with CC, it may still occur.

Hughes and colleagues published a systematic review evaluating CC (daily doses of 10-250mg po for 5 days) for induction of ovulation in women with oligomenorrhea (WHO type II anovulation including women with PCOS). Outcomes assessed were pregnancy and ovulation. Four cross over RCTs were included in the quantitative summary. CC was more effective than placebo for both ovulation and achievement of pregnancy at doses of 50 –250 mg per day. ( OR 6.82, [CI: 3.92 – 11.85] and OR 3.41, [CI: 4.23 – 9.48] respectively).(22)

Gonadotropin therapy (controlled ovarian hyperstimulation) with in-utero insemination (COH- IUI) is also very effective in this patient group. A conservative stimulation using low dose gonadotropins in a step up protocol is most appropriate. However, these patients are at high risk for ovarian hyperstimulation syndrome, multiple pregnancy and very brittle responses requiring frequent ultrasound and blood monitoring. (57)

Conventional stimulation techniques for PCOS involved CC or COH-IUI. The utilization of in-vitro fertilization for achievement of pregnancy in patients with PCOS was first

introduced in those patients resistant to CC and or gonadotropin therapy.(58, 59)  
Generally, IVF is reserved for cases of PCOS that are either extremely brittle or refractory to conventional therapy. IVF may be of benefit in avoiding multiple gestation and ovarian hyperstimulation syndrome. Some authors have argued whether IVF outcome is comparable to COH-IUI. Concerns about oocyte maturity, quality, and implantation have been raised. To date, no RCTs appear to exist comparing COH-IUI to IVF in this patient group. Prospective cohorts comparing COH-IVF do not appear to show a difference. However, retrospective cohorts suggest IVF may compromise outcome in this group but selection bias may be a contributing factor. (60-62)

Metformin is a biguanide insulin sensitizer. It is labeled as class B in pregnancy. Sufficient human data is not available but the drug has not been associated with congenital defects in animals. It has been used in the treatment of women with diabetes mellitus II without negative effects on the fetus. Recent, small series of women who have taken metformin throughout pregnancy to prevent miscarriage and gestational diabetes have not reported congenital defects. The mechanism of action is mainly the inhibition of hepatic gluconeogenesis. Metformin also increases peripheral glucose utilization and insulin sensitivity, but it is not associated with hypoglycemia.(63)

Theoretically, metformin reduces insulin response by decreasing hepatic gluconeogenesis and reducing androgen levels, which allow resumption of normal menstrual cyclicity. Metformin should not confer the same risks of ovarian hyperstimulation and multiple pregnancy as clomiphene citrate since metformin returns patients to spontaneous ovulation not superovulation. Also, metformin should not have the same negative effects on the cervical mucous and endometrium as clomiphene citrate. Metformin was first suggested as a treatment for ovulation induction in women with PCOS in the early 1990's. The first reports demonstrated resumption of menstrual cyclicity in 21 of 22 patients with PCOS.(64) Since then many studies have been conducted on small samples of women with PCOS with the primary outcomes ranging from alteration of metabolic profiles to reinstatement of ovulation.(65) Most of these studies were observational and many retrospective. Several studies have demonstrated positive effects on the surrogate

outcomes of metabolites such as androgen level and fasting insulin levels as well as ovulation.(66) Also, none of the studies that do exist use pregnancy as a primary outcome.

Patients with PCOS appear have an increased risk of recurrent miscarriage. The explanation for this is still unclear. Some authors hypothesize that increased LH levels, elevated circulating androgen levels, and / or hyperinsulinemia induce a hostile environment for the fetus. Recently PAI-1 (plasminogen activating factor) has been suggested to be responsible for hypofibrinolysis in these patients. Even more recently, some authors have suggested that treatment with insulin sensitizers may improve the miscarriage rate. (67, 68)

An, as yet, unpublished randomized controlled trial of metformin plus IVF versus IVF alone demonstrated a trend towards increased implantation and biochemical pregnancy rates but not clinical pregnancy rates. Metformin, in this group, was stopped on the day of human chorionic gonadotropin (hCG), before embryo transfer. The regression in pregnancies rates may be due to a early loss rate after discontinuation of metformin, if in fact metformin is protective against recurrent pregnancy loss. Although the sample size of this study was small ( N = 73) and the power inappropriately calculated for multiple outcomes, there was a trend ( $p < 0.05$ ). (69, 70)

#### **2.1.1.4 Rationale for proposed study**

Since Velazquez's original paper, several studies have been conducted on small samples of women with PCOS with the primary outcomes ranging from alteration of metabolic profiles to reinstatement of ovulation. Most of these studies were observational and many retrospective. Some studies of metabolic parameters investigated metformin versus placebo. The few randomized controlled trials investigating metformin for induction of ovulation did so in women who had already failed CC treatment. These studies compared metformin + CC to placebo +CC. There are no direct, head to head RCTs of metformin versus CC in these patients. This is despite the fact that several studies have demonstrated positive effects on the surrogate outcomes of metabolites such as androgen level and fasting insulin levels as well as ovulation. Also, none of the studies that do exist use

pregnancy as a primary outcome. The systematic review we are undertaking will provide a more comprehensive assessment of the available literature.

The significance of this proposed study is great. Metformin, in association with clomiphene citrate, has been shown, in small sample sizes to be effective at modifying the surrogate endpoints associated with induction of ovulation. There has never been a direct comparison of metformin with clomiphene citrate by a randomized controlled trial; the systematic review we are undertaking will provide a comprehensive assessment of the literature. The need for such a trial has been recognized. Metformin has been shown to be useful for resuming normal menstrual cyclicity in patients with PCOS. The objective of metformin therapy by itself would be to initiate normal ovulation as opposed to superovulation. The negative estrogen effect on the endometrium and cervical mucus associated with clomiphene citrate should not accompany metformin therapy. Also, since superovulation is not induced, less monitoring would be required, and insemination should not be necessary since cervical mucus should not be compromised. Metformin is less expensive than other methods of ovulation induction such as CC or gonadotropins and would not be associated with the risks of hyperstimulation and multiple pregnancy. It could be a first line therapy for such patients and could be carried out in the family doctor's office. CC therapy is limited to a maximum of 12 months because a nested case-control study suggested it might be associated with an increased risk of ovarian cancer. (21) Although the study design was poor, this dogma is still accepted in clinical practice. Many PCOS patients will be resistant to CC therapy and treatment with metformin prior to CC therapy may be more efficient and avoid unnecessary exposure to CC.

### **2.1.2 Objectives**

#### **1) Systematic review:**

- a) To conduct a systematic review, using Cochrane, MOOSE, and QUORUM guidelines ( RCTs and cohort studies), and provide a quantitative summary where applicable regarding the following question.



Question: “Is metformin of benefit compared to clomiphene citrate for induction of ovulation and pregnancy rates in patients with PCOS?”

**2) Randomized controlled trial:**

- a) To design and implement a randomized controlled trial following CIHR guidelines to answer the following study question.

Question: In patients who have a clinical diagnosis of PCOS, is metformin of benefit compared to clomiphene citrate for 1) induction of ovulation and 2) achievement of pregnancy? The primary objective is to test the efficacy of metformin versus clomiphene citrate on the primary outcome which is pregnancy. The secondary objective is to compare metformin to clomiphene citrate on the secondary outcome which is ovulation.

## **2.2 PCOS Systematic Review of Metformin vs CC**

### **2.2.1 Methods for Conducting Systematic Review**

#### **2.2.1.1 Search Strategy**

A **computerized literature** search of the following databases, using an OVID vendor and Polaris interface, was conducted: Medline, Premedline, Current Contents, Biological abstracts, and EMBASE. The databases were searched for the last 25 years. For databases without MeSH headings, textwords were used. Adjacency operators, and truncation were employed. (Table 2) A preliminary search had already been conducted to maximize potential keywords. We did not apply a study filter and the search was not limited by language or year of publication. It was repeated at two-week intervals until the completion of the meta-analysis at the end of September, 2003.

A **hand search** of ten years of the following journals was also completed: Fertility and Sterility, Human Reproduction, Journal of Clinical Endocrinology, New England Journal of Medicine, the Lancet, and the American Journal of Epidemiology. Online indexing facilitated this process. These journals were chosen on the basis of speciality and previously published trials in this clinical area.

To avoid publication bias, we searched for **grey literature**. Three content experts were contacted. Also, 10 years of conference proceedings for the following were searched: The Canadian Fertility and Andrology Society, The American Society for Reproductive Medicine, and the Society for Gynecologic Investigation. These conferences were chosen on the basis of speciality.

The Cochrane database was searched for systematic reviews, study protocols, or completed, registered clinical trials. Titles and abstracts were screened and articles retrieved if they passed the relevance filter or if there was uncertainty whether or not they were relevant. Bibliographies of review articles, systematic reviews, and retrieved articles will also be searched for candidate articles. Retrieved articles were reviewed for inclusion/exclusion criteria and those articles meeting these criteria were kept for critical appraisal and data collection.

**Table 2: PCOS systematic review search strategy**

<b><u>A. Key words</u></b>	
<b>Study population</b>	<b>Intervention</b>
PCOS or PCOS	Clomid
Polycystic adj ovar\$ adj (syndrome or disease)	Clomiphene adj citrate
Stein adj leventhal adj syndrome	Metformin
anovulat\$	Ovulation adj induc\$
Amenorrhea	Serophene
Hyperandrogen\$	Biguanides
Hyperinsulim\$	Controlled adj ovar\$ adj hyperstimula\$
	Insemina\$
	COH
	SO-IUI
<b><u>B. Databases</u></b>	<b><u>C. Hand search of journals and conference proceedings*</u></b>
Cochrane controlled trials register	Fertility and Sterility
Cochrane database of systematic reviews	Human Reproduction
EMBASE ( 1980 – 2002)	New England Journal of Medicine
Premedline	Journal of Clinical Endocrinology and Metabolism
Medline (10 years)	Lancet
Current Contents	Journal of the American Medical Association
Biological abstracts	Cochrane Database
	American Society of Reproductive Medicine
	*facilitated by online indexing

### **2.2.1.2 Inclusion/ Exclusion**

One reviewer (SK) completed all stages of the review selection process. MOOSE and QUOROM guidelines were followed.(6, 7) Titles and abstracts were screened and articles retrieved if they passed the relevance filter (section 2.4.1)or if there was uncertainty as to whether or not they were relevant. Bibliographies of review articles, systematic reviews, and retrieved studies were also searched for candidate articles.(71) Retrieved articles were then reviewed for inclusion/ exclusion criteria. Those articles that met the criteria were then kept for critical appraisal and data collection. The reviewer was not blinded at any point to the authors or sources of publication as the evidence for such blinding is weak and the reviewer was likely to be familiar with the literature. (72)

### **2.2.1.3 Quality Assessment**

The Jadad scale was used for quality assessment of randomized-controlled trials.(73) Quality assessment involved evaluation of patient selection, assessment of exposure or outcomes, administration of interventions, and controls for confounding factors. (Appendix A) The Ottawa NewCastle Scale for Observational Studies (NOS) was to be used for quality assessment of observational studies.

### **2.2.1.4 Study Inclusion (PICOS)**

“PICOS” defines the inclusion criteria for a systematic review. The acronym represents: types of **P**articipants/**P**opulation, types of **I**ntervention, types of **C**omparison, **O**utcomes and types of **S**tudies.

**Study population:** The population of interest was women of reproductive age (age 20 to 45) who have a history of PCOS as defined by chronic oligo-ovulation/anovulation, chronic hyperandrogenism, and infertility. Other diagnosis for infertility such as male factor and tubal infertility were excluded.

**Intervention and comparison:** The drugs of interest were clomiphene citrate (doses 50 – 150 mg) and metformin and the comparisons were as follows: clomiphene citrate ( in

doses of 50 to 150 mg) versus metformin; clomiphene citrate versus clomiphene citrate + metformin; metformin + clomiphene versus placebo + clomiphene citrate; metformin versus placebo.

**Outcome:** The outcomes of interest were pregnancy (primary outcome) and ovulation (secondary outcome).

**Study design:** Both randomized –controlled trials and controlled cohort studies were evaluated. A quantitative summary was done where appropriate for homogeneous studies but RCTS were evaluated separately from cohort studies.

### **2.2.1.5 Quantitative Methods**

Where appropriate, Revman 4.1 and Metaview 4.0 were used to analyze data. A relative risk estimate, with confidence intervals, was calculated from the combined results from randomized controlled trials. We analyzed the data using both a fixed effects model and random effects model. If interstudy variation was present and/or different levels of treatment were used, the random effects model would be presented. If not, the fixed effects model would be presented. A chi-square test was done to determine the significance of the association for each of the outcomes evaluated and heterogeneity was determined by the Cochran Q test in order to demonstrate statistical comparability of the studies. These results are illustrated graphically in the form of Forrest plots. Funnel plots have been constructed to represent the likelihood of publication bias. The precision of the search strategy was determined by the formula of Normand. Recall of the search strategy was also calculated.(74) Where appropriate, sensitivity analysis and subgroup analysis was done to determine the significance of contributing factors to the overall results.

## **2.2.2 Results of PCOS Systematic Review**

### **2.2.2.1 Study inclusion**

A total of 1057 citations were selected for review which was reduced to 945 citations when limited to human females. Letters, review articles, and subjects clearly unrelated to

our PICOS statement were rejected. Articles whose relevance was unclear were retrieved after removal of duplicates. Removal of duplicates was manually checked. Three additional abstracts were found through a hand search of relevant journals and conference proceedings. The relevance filter was applied to the titles and abstracts of 668 articles. This resulted in 37 articles as potential subjects for analysis. All of these articles were retrieved. Inclusion/exclusion criteria were then applied, as in Figure 2, and 8 papers (7 published RCTS and 1 abstract of an RCT) were included in the final analysis. The specific steps in the selection of the studies are given in Figure 2. Overall precision of the search strategy was 0.76% (8/1057). Recall was 88% (7/8). No randomized controlled trials were excluded for multiple publication.

All included RCTs were published between 1998 and 2002.(66, 75-81) Although we did not limit the language of publication, all retrieved articles were in English.

Both observational and randomized controlled trials were included in the study search. Controlled cohort studies, not case-series or uncontrolled cohorts, were to be included in the analysis. However, no appropriate controlled cohort studies were found. Therefore, the analysis from this point forward includes only RCTs. The Jadad quality scale for randomized controlled trials was applied to all included studies and is shown in Table 3.(82)

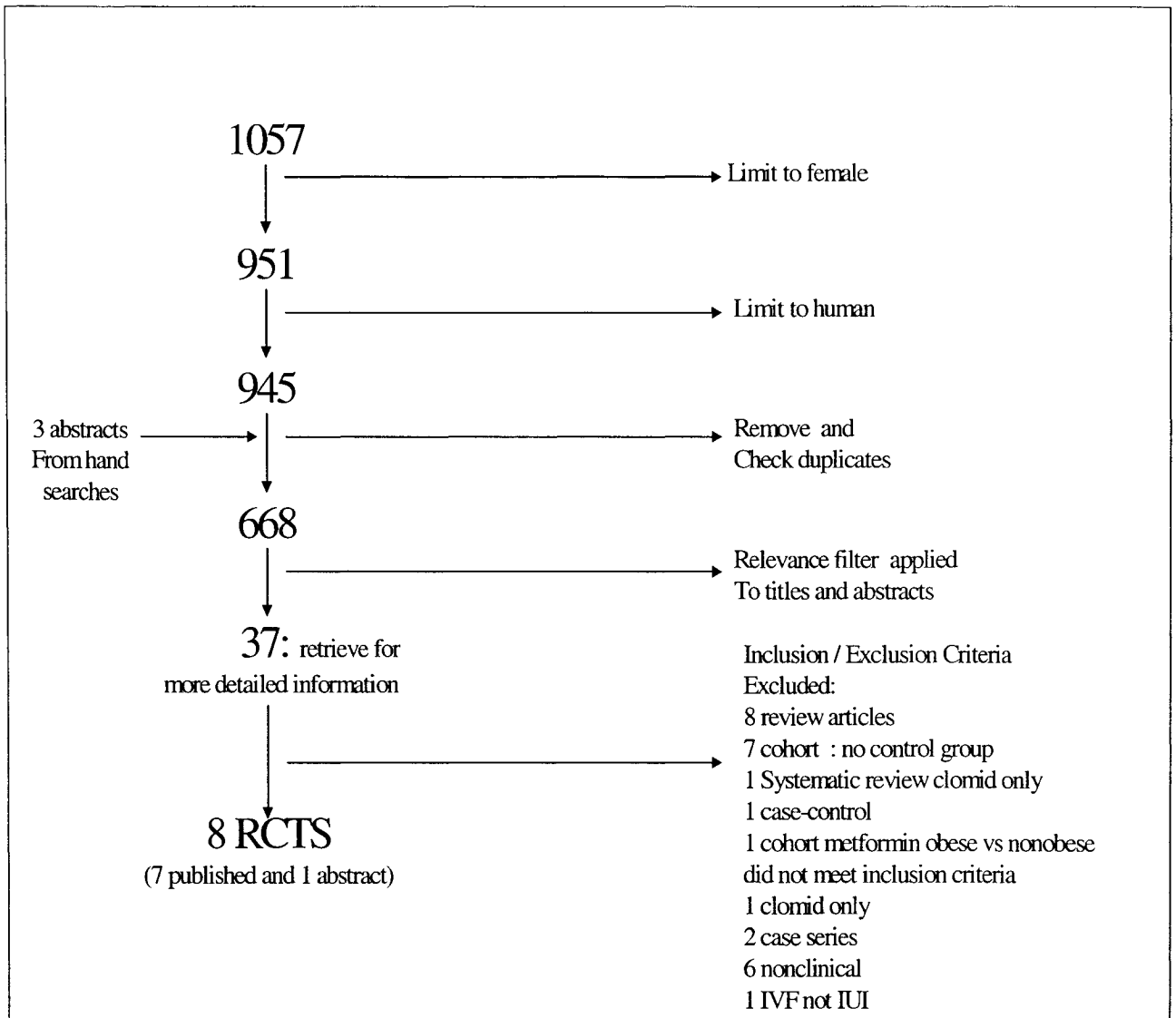
#### **2.2.1.2 Characteristics of Included Studies**

Studies were assessed for comparability of study population by age, fertility, and method of diagnosis of PCOS. We also reviewed investigation for potential confounders, compliance, and contamination or cross-over. None of the studies commented on contamination or cross-over; only one study mentioned compliance but that study did not describe the methods used to assess compliance.(81)

All studies, except Singh et al., investigated metformin versus placebo as the primary intervention. All studies except 3 used pregnancy as an outcome but only 1 abstract assessed pregnancy as a primary outcome. (66, 77-79) All studies except 3 had infertile,

PCOS women as their study population.(66, 75, 77) Several studies tried to compare metformin plus CC to CC alone or with placebo.(76-78, 81) However, all of these studies did so in a sequential manner in those patients who failed to conceive or ovulate with metformin. Patients in those studies were initially randomized to metformin versus placebo and if they did not ovulate proceeded into an open label extension where they continued with the initial treatment (metformin versus placebo) but added CC. The duration of intervention and follow-up varies from 1 – 6 months. The sizes of the studies varied from 18 patients to 100 patients. Each of the studies included oligomenorrhea in the criteria for diagnosis of PCOS, four studies included hyperandrogenemia, and three studies included ultrasound criteria. Four studies included participants who had previously failed CC treatment or were deemed CC resistant. The randomization method was explicit in only three studies. The details of each study are provided in Table 4.

**Figure 2: Study inclusion**





**Table 3: Quality assessment by Jadad scale of PCOS RCT**

<b>Study</b>	<b>Randomized?</b>	<b>Randomization Method Described</b>	<b>Double Blind?</b>	<b>Double Blind Method Described</b>	<b>Withdrawals &amp; Dropouts Described?</b>	<b>Score *</b>
		<b>Appropriate (+1) Inappropriate (-1)</b>		<b>Appropriate (+1) Inappropriate (-1)</b>		
Kocak, 2002*	+1	-1	+1	+1	+1	3
Fleming, 2002*	+1	+1	+1	+1	+1	5
Sturrock, 2001*	+1		+1			2
Vandermolen, 2001*	+1	+1	+1	+1	+1	5
Ng, 2001*	+1	+1	+1	+1		4
Singh, 2001	+1					1
Moggetti, 2000*	+1		+1	+1		3
Nestler, 1998*	+1					1

\*Scores range from 0-5 with the higher scores indicating the better quality studies

**Table 4** Characteristics of randomized controlled trials

Study and intervention	No. of subjects: a. Study; b. Control	Ascertainment of outcome: A. Ovulation B. Pregnancy	Location	Intervention	Method of randomization	Outcome	Sources of bias
Kocak <i>et al.</i> (2002)	a. 23 b. 23	A, B	Turkey	cycle 1: metformin versus placebo; cycle 2: metformin + CC versus placebo + CC But hCG at US > 18mm 14 weeks metformin versus placebo	sequential odd numbers metformin; even numbers placebo	ovulation n/a placebo versus metformin; pregnancy: metformin 1/23 versus placebo 0/23 hCG given in 2nd arm ovulation: metformin 37/45 versus placebo 30/47	no comment  no comment
Fleming <i>et al.</i> (2002)	a. 45 b. 47	A, B	UK	metformin versus placebo × 12 weeks metformin + CC versus placebo + CC × 1 month	computer-generated blocks of 4 sample size 2n = 310 intention to treat not detailed	ovulation: metformin 0/12 versus placebo 1/14 pregnancy: metformin + CC 3/12 versus placebo + CC 2/14	no comment
Sturrock <i>et al.</i> (2001)	a. 12 b. 14	A, B	UK	metformin versus placebo × 7 weeks metformin + CC versus placebo + CC × max 6 cycles	computer-generated blocks of six intention to treat double-blind placebo	ovulation: metformin 1/12 versus placebo 1/15 metformin + CC 8/11 versus placebo + CC 3/14 pregnancy: metformin 0 versus placebo 0 metformin + CC 6/11 versus placebo + CC 1/14	no comment
Vandermolen <i>et al.</i> (2001)	a. 11 b. 15	A	USA	metformin versus placebo × 3 months metformin + CC versus placebo + CC × 1 cycle	computer-generated double-blind placebo sample size 2n = 16 based on Nestler's data	ovulation: metformin 3/9 versus placebo 3/9 metformin + CC 2/4 versus placebo + CC 4/4 pregnancy: met 1/9 versus placebo 2/9 metformin + CC 2/4 versus placebo + CC 4/4	no comment
Ng <i>et al.</i> (2001)	a. 9 b. 9	B	China	metformin + CC versus CC × 4 months	method not described	pregnancy: metformin + CC 31/53 versus cc 4/47	no comment
Singh <i>et al.</i> (2001)	a. 53 b. 47	B	India	6 months metformin versus placebo metformin versus placebo × 5 weeks, then metformin + CC versus placebo + CC	method not described double-blind placebo method not described	ovulation: metformin 5/11 versus placebo 0/11 improved menses ovulation: metformin 12/35 versus placebo 1/26 placebo + CC 2/25 versus metformin + CC 19/21	no comment  no comment
Moghetti <i>et al.</i> (2000)	a. 11 b. 11	A	Italy				no comment
Nestler <i>et al.</i> (1998)	a. 35 b. 26	A Progesterone > 8 mg/ml	USA				no comment

Table 4 Characteristics of randomized controlled trials (continued)

Study	Study group: a. PCOS b. Infertile PCOS	1. Primary 2. Secondary a. Ovulation b. pregnancy	Criteria for diagnosis of PCOS	Confounders: a. Duration infertility b. Previous treatment c. Age d. Body mass index e. Hyperandrogenaemia f. Hyperinsulinaemia	Compliance assessed	Length of follow-up (no. of cycles)	Assessment for tubal or male factor	Assessment for contamination/ cross-over
Kocak <i>et al.</i> (2002)	B	A, B	oligomenorrhoea hyperandrogenemia multiple subcapsular follicles on ultrasound	b, c, d, e, f	no comment	cycle 1: metformin versus placebo cycle 2: metformin CC > versus placebo + CC but hCG at CC > 18 14 weeks metformin versus placebo	tubal and male factor ruled out	no comment
Fleming <i>et al.</i> (2002)	A, B* age < 35 years	A, B*	oligo/amenorrhoea Adams <i>et al.</i> ultrasound criteria	c, d, e, f	drop-outs reported but compliance assessment not described SS calculation no comment SS calculation	6 months metformin versus placebo	not assessed/ pts not desiring pregnancy advised to use barrier contraception no comment	no comment
Sturrock <i>et al.</i> (2001)	B. 18–40	A, B	oligomenorrhoeic	b Assess insulin, androgens but no comparison c, d, e, f	1 patient excluded because 'non- compliance' compliance measures not described no comment	metformin versus placebo × 7 weeks metformin + CC versus placebo + CC × max six cycles	tubal & male factor ruled out	no comment
Vandermolen <i>et al.</i> (2001)	B age 18–35 years	A, B	oligoovulation hyperandrogenaemia	a, b, c, d, e, f	no comment	metformin versus placebo × 3 months metformin + CC versus placebo + CC × 1 cycle	tubal and male factor ruled out	no comment
Ng <i>et al.</i> (2001)	B. age < 40 years	A	irregular cycles and anovulation' by midluteal progesterone CC resistant	a, b, c, d, e, f	no comment	metformin + CC versus CC × 4 months	no comment	no comment
Singh <i>et al.</i> (2001)	B	B	oligo/amenorrhoea 'ultrasound' LH/FSH > 2	c	no comment	6 months metformin versus placebo	not assessed	no comment
Moghetti <i>et al.</i> (2000)	A. age 18–35 years	A	oligomenorrhoea hyperandrogenaemia normal GTT BMI > 28 kg/m <sup>2</sup>	c BMI > placebo group androgen > metformin group b. 31 or 61 patients previously received CC c, d, e, f	no comment	metformin versus placebo × 5 weeks. then metformin + CC versus placebo + CC	abstinence/barrier contraception advised	no comment
Nestler <i>et al.</i> (1998)	A	A	oligomenorrhoea hyperandrogenaemia ultrasound normal GTT		no comment			no comment

CC = cloniphene citrate; GTT = glucose tolerance test; BMI = body mass index.

### 2.2.2.3 Meta-analysis

A total of 8 studies were included in the qualitative analysis (7 published RCTs and 1 abstract) but only the 7 RCTs were included in the final quantitative analysis since the numbers in the abstract were not explicit enough for inclusion.

Figures 3 to 5 summarize the results in the form of Forrest plots and are further detailed below. Figure 3 compares metformin to placebo in infertile PCOS patients for the outcomes ovulation and pregnancy. Figure 4 compares studies which included a longer duration of metformin versus placebo therapy for the outcome ovulation and also investigates metformin versus placebo therapy in PCOS patients who may not have been infertile. Figure 5 investigates the results of randomized controlled trials which randomized metformin to placebo but then continued a sequential, open label comparison of metformin plus CC versus placebo plus CC of patients in each respective arm who did not ovulate.

Although the Cochran Q test did not suggest statistical heterogeneity in each of the following comparisons, there are several potential clinical sources of heterogeneity. Of particular concern were the following issues. None of the studies evaluated insulin resistance prior to administration of metformin. In most studies, the clinical picture of obesity as a characteristic of most patients with PCOS was considered sufficient evidence of insulin resistance. However, the study by Ng and colleagues evaluated Chinese women who met ultrasound criteria of PCOS (Adam's criteria), who were anovulatory and CC resistant but who were not obese (Average BMI 23.8-24.1 kg/m<sup>2</sup>). Duration of follow-up or time to outcome also varied from 34 days to 16 weeks as did the duration of treatment with metformin (28 days to 3 months) before addition of CC. Fortunately, all studies used hard, biochemical evidence of outcome for both ovulation and pregnancy. All studies also detailed their used of consistent doses for the medications being investigated. Flemming and Moghetti also investigated women who were not necessarily infertile but who suffered other manifestations of PCOS such as cycle irregularity. This group was investigated separately and may not have been similarly motivated to comply with the protocols.

Five studies were included in the first quantitative summary of metformin versus placebo for ovulation induction in patients with PCOS and these results summarize the outcome of 222 patients. (Figure 3, panel 1) While there did not appear to be heterogeneity, either by the fixed or random effects model, the forest plot clearly shows that the first study by Fleming et al. weighs highest, contributing 92 patients, has the most narrow confidence interval and largest influence on the results. The Nestler study contributes 60 patients and has a point estimate far to the right with a significant, although wide confidence interval. The point estimates for the 3 smaller studies are equally and equivocally dispersed, with wide, nonsignificant confidence intervals. Fixed effects models are presented for the analysis. However, both fixed and random effects analyses had been performed since there was concern that the study populations and interventions may not have been completely homogeneous. Figure 3, panel 1 and Figure 4, panel 1 indicate that metformin is 50% better than placebo for ovulation induction in infertile PCOS patients (RR 1.50) but this benefit is not necessarily improved with longer duration of therapy (> 3 months – RR 1.37). The Fleming study followed patients on metformin for 16 weeks and heavily influences both outcomes.

Two studies also investigated metformin for cycle regulation. Metformin is also of benefit in noninfertile PCOS patients for cycle regulation compared to placebo. (Figure 4, panel 2, RR 1.45, [CI 1.11, 1.90])

Four studies (Figure 3, panel 2) investigated metformin versus placebo for achievement of pregnancy. Again the follow-up time to pregnancy was short and varied from 28 days to 7 weeks. The study by Kocak and colleagues was relatively large with 46 patients; the other 3 studies each had less than 20 patients. To date metformin is not of confirmed benefit versus placebo for achievement of pregnancy for this duration of follow-up. (Figure 3, panel 2, RR 1.07, [CI 0.20, 5.74]) This may be due to the fact that most studies used ovulation as their primary outcome, those studies that assessed pregnancy had a short follow-up time to pregnancy, and most studies had small sample sizes.

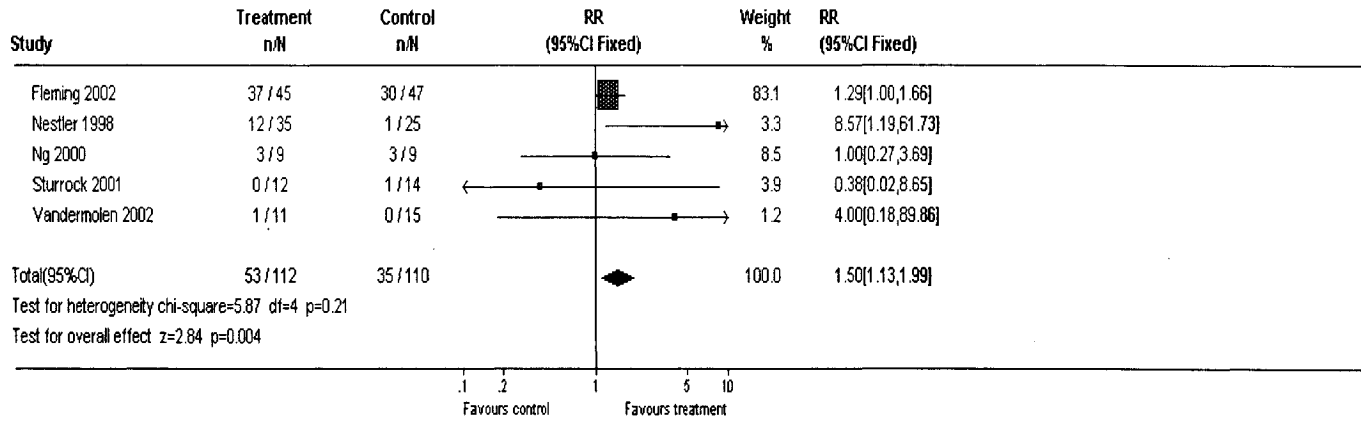
The analysis of metformin plus CC versus CC alone or with placebo is represented here. The study design for this component is reported as a cross-over design but is actually a RCT with an extended, open label sequential arms for each of the randomized arms. Subjects were randomized to either metformin or placebo. Those subjects who did not ovulate continued in the same randomized arm (metformin or placebo) and added CC. Some of these studies reported the second arm as a “cross-over” design. This is incorrect. Firstly, the nature of infertility therapy makes cross-over trials an inappropriate design. Secondly, in order to conduct a cross-over trial, the two therapies being compared must be “washed out” within a short window of time between arms, the outcome must also be completely reversible within this window so that the patient returns to baseline, and the chance of successful outcome should not alter with time such that one therapy has a disadvantage over the other by being administered in the later arm or that the chance of success varies with time. Nevertheless, we did do a quantitative analysis on the data supplied to provide some assessment of the above interventions for ovulation and pregnancy, since this is a very common clinical question and this is the best available evidence to date. (Figure 5) Four studies investigated the above design for ovulation in PCOS patients and each of these studies was reasonably, similarly weighted. Two studies investigated the stated study design for pregnancy outcome. The Sturrock study had the most weight and a RR of 1.75 but the confidence interval crossed 1. The Vandermolen study had a RR of 7.64 and a significant confidence interval. Given the stated limitations of such a study design, we found that metformin plus CC may be superior to CC alone or with placebo with regards to ovulation (RR 3.04, [CI 1.77, 5.24]) and pregnancy (RR 3.65, [CI 1.11, 11.99]).

Funnel plots for each comparison did not suggest obvious publication bias but only 2 to 5 points were available for each analysis. (Figures 6-8)

**Figure 3: Metformin vs placebo in infertile PCOS patients; Outcome 01 Ovulation induction; Outcome 02 Pregnancy achievement**

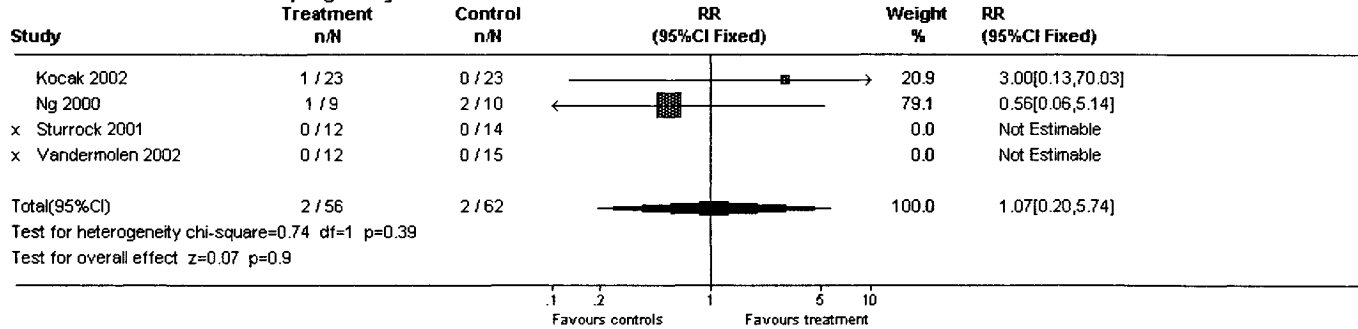
Comparison: 01 Metformin versus placebo in infertile PCOS pts

Outcome: 01 outcome = ovulation



Comparison: 01 Metformin versus placebo in infertile PCOS pts

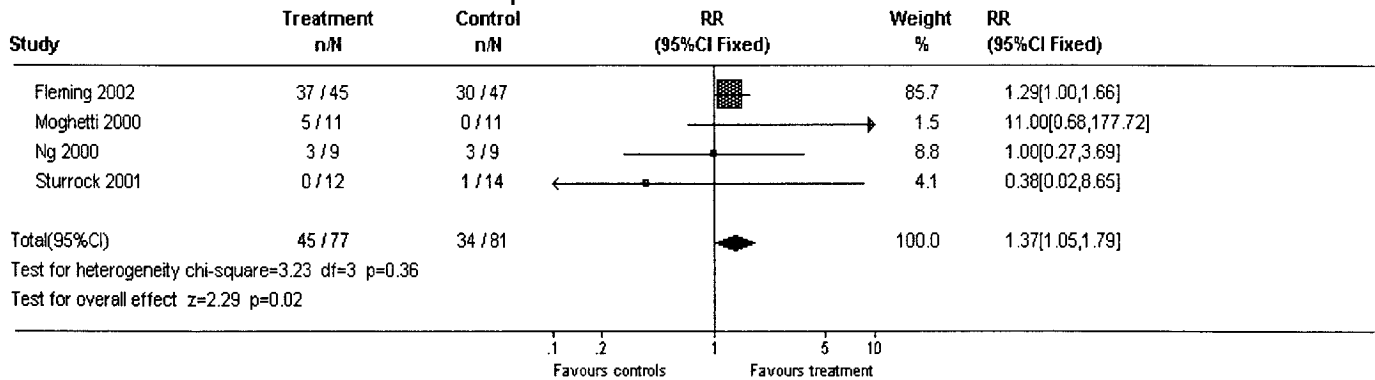
Outcome: 02 outcome = pregnancy



**Figure 4: Metformin vs placebo for cycle regulation in PCOS patients**  
**Outcome 01 Ovulation for duration of therapy > 3 months ; Outcome 02 Ovulation (cycle regulation) for patients who are not complaining of infertility**

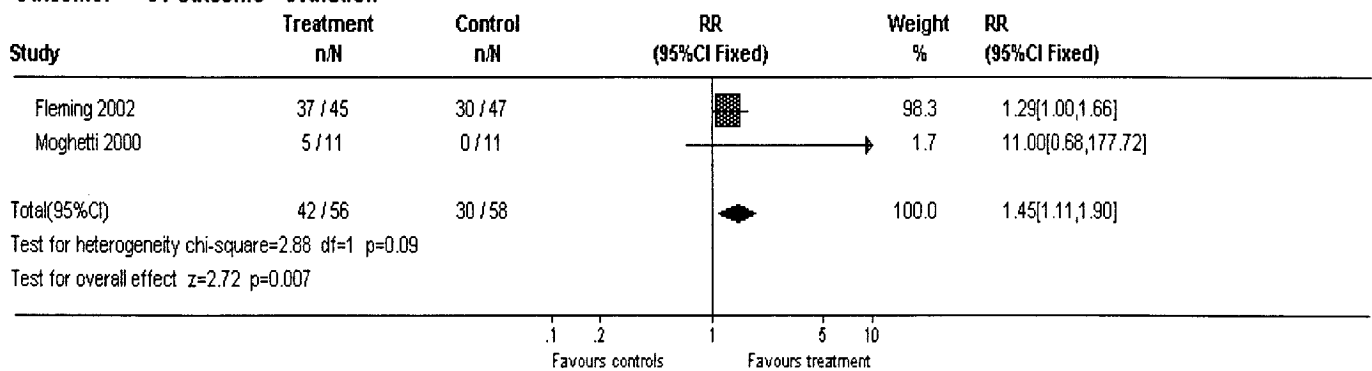
**Comparison: 01 Metformin versus placebo in infertile PCOS pts**

**Outcome: 03 outcome ovulation in RCT >3mos placebo vs metformin**



**Comparison: 02 Metformin versus placebo in PCOS ( not nec infertile) pts**

**Outcome: 01 outcome - ovulation**

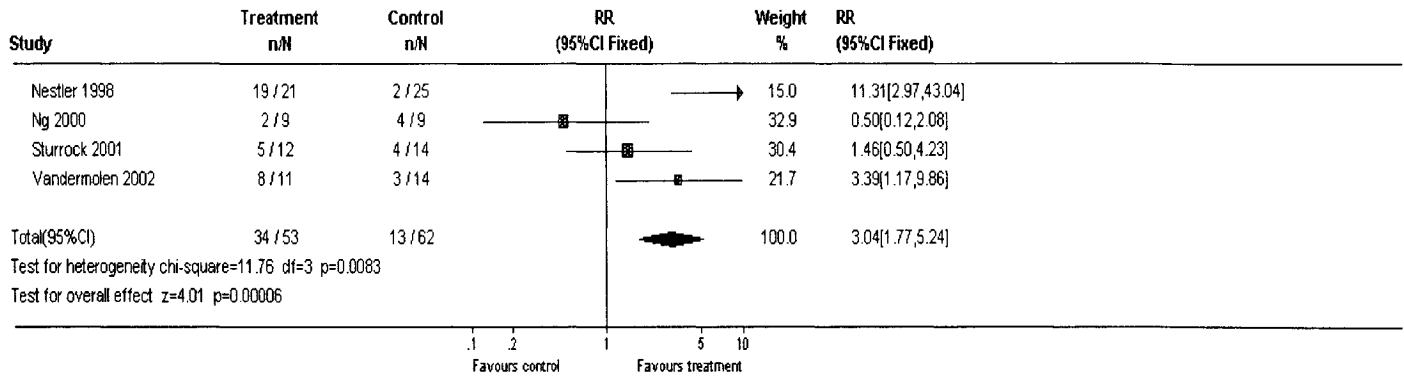




**Figure 5: Metformin plus clomiphene citrate vs placebo plus clomiphene citrate;  
01 Ovulation induction ; 02 Pregnancy achievement**

Comparison: 03 metformin + cc vs placebo + cc

Outcome: 01 ovulation



Comparison: 03 metformin + cc vs placebo + cc

Outcome: 02 pregnancy

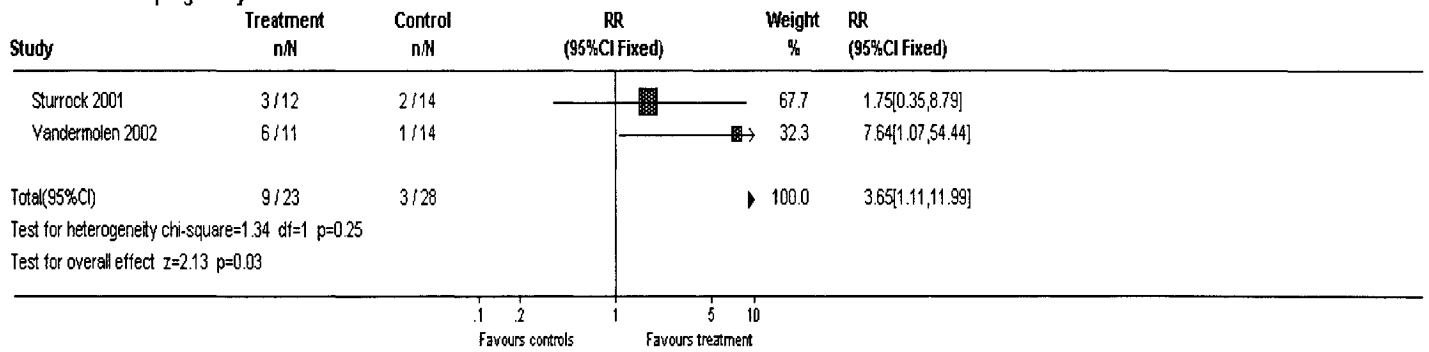
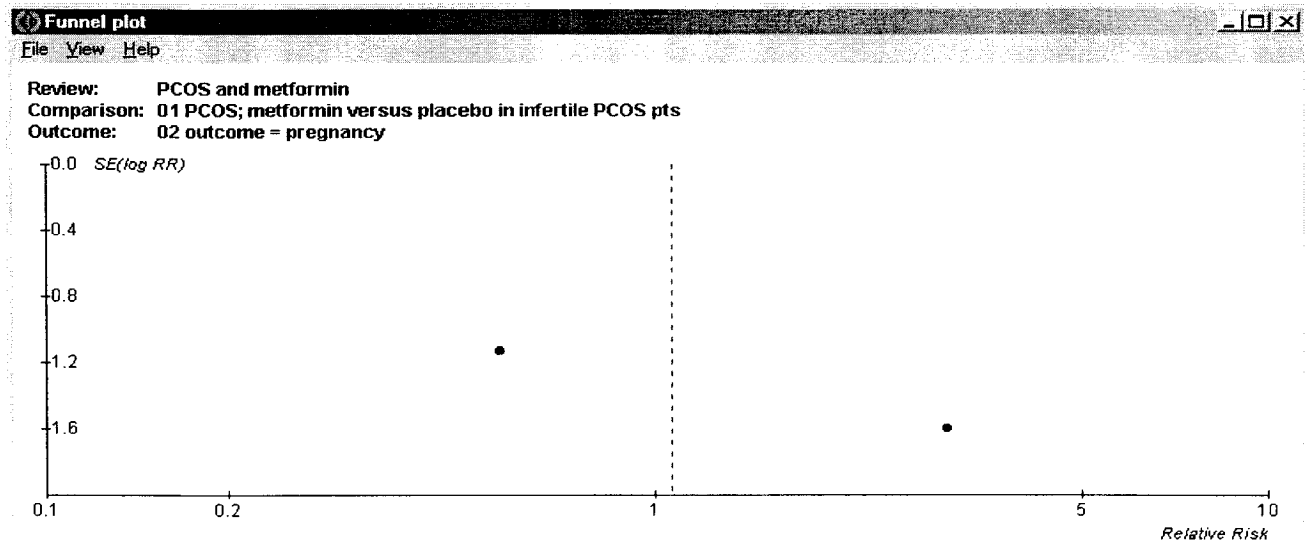
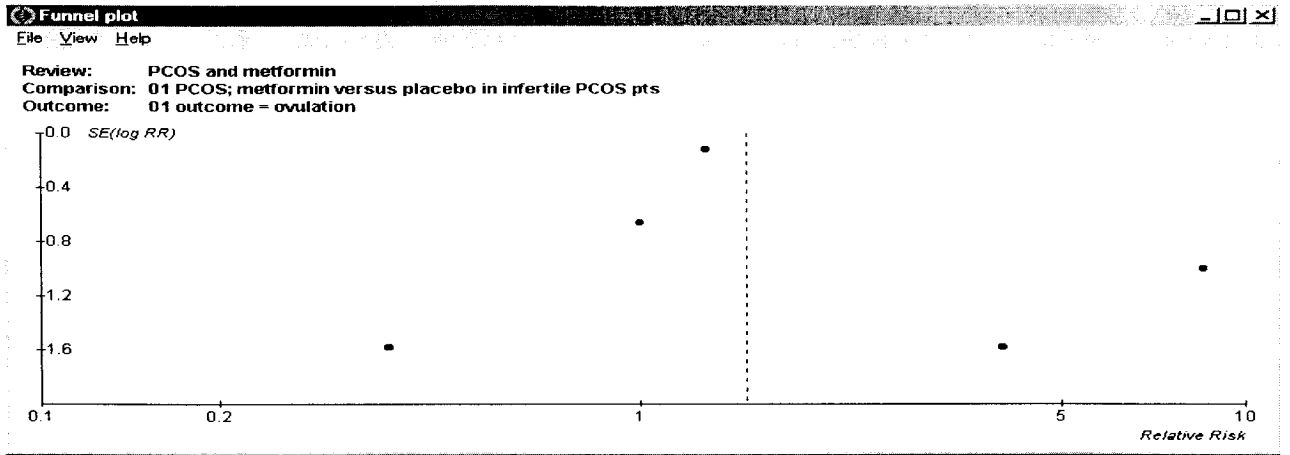
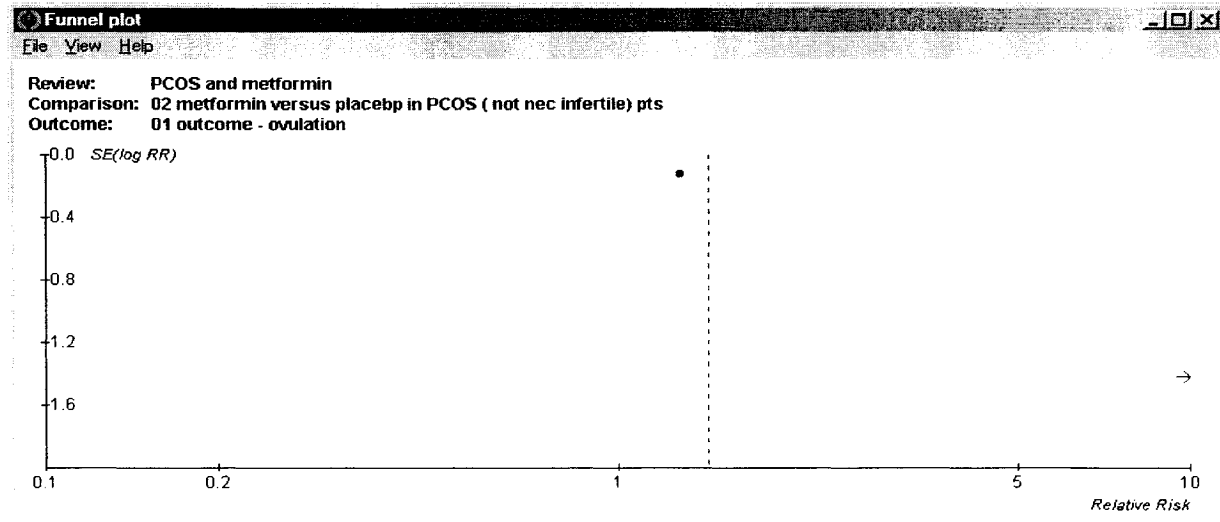
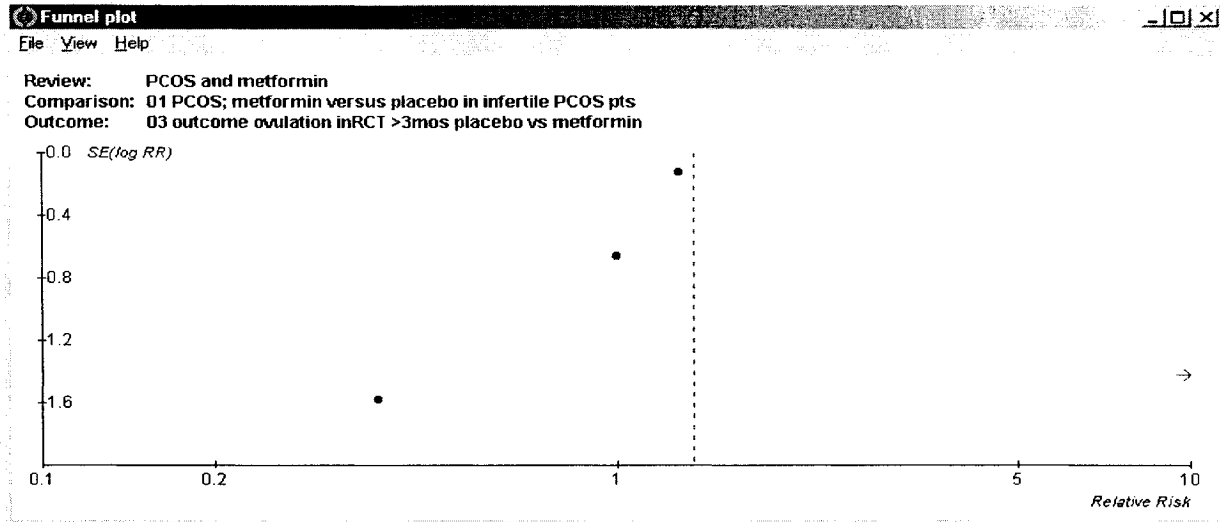


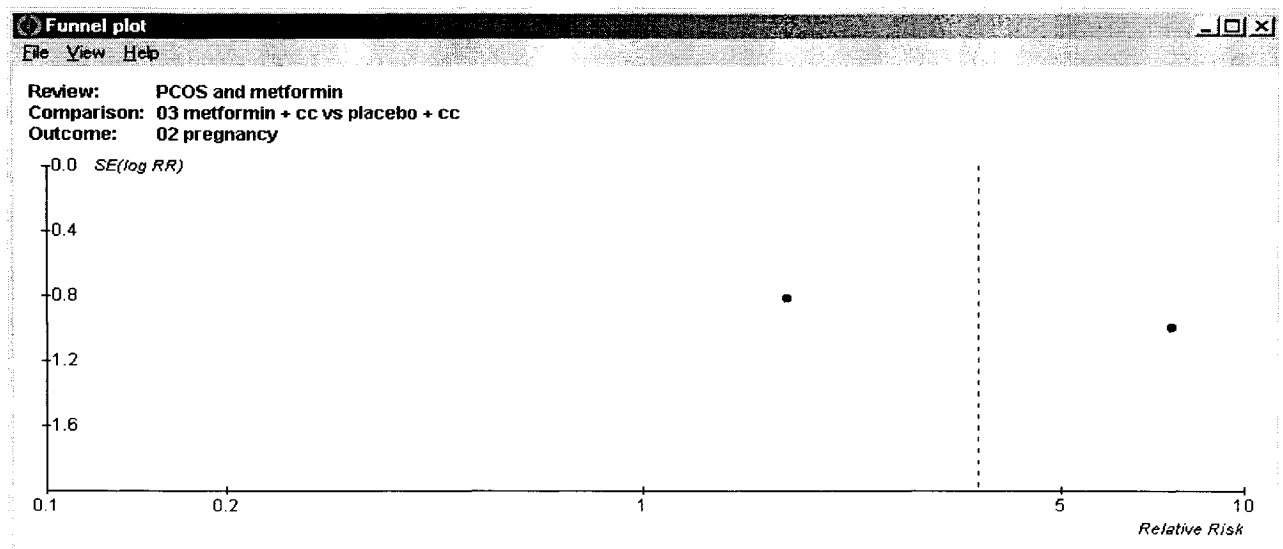
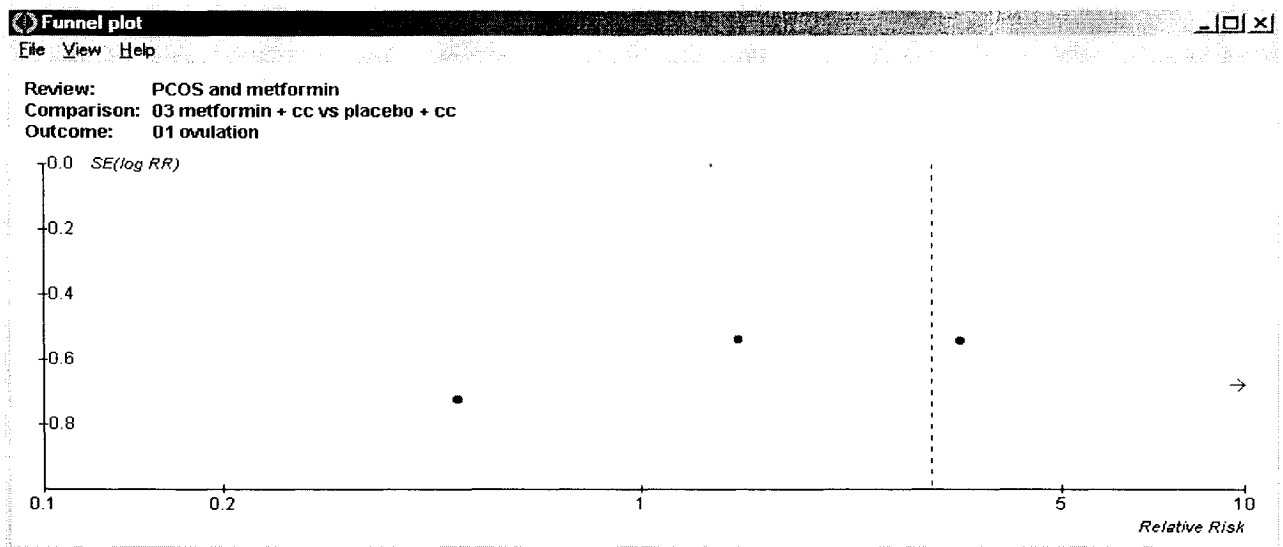
Figure 6: Funnel plots of studies investigating metformin vs placebo in PCOS patients; Outcome 01 Ovulation induction; Outcome 02 Pregnancy achievement



**Figure 7: Funnel plots of studies investigating metformin vs placebo in PCOS patients; Outcome 01: Ovulation for duration of therapy > 3 months; Outcome 02: Ovulation (cycle regulation) for patients who are not complaining of infertility**



**Figure 8: Funnel plots of metformin + CC vs placebo + CC in PCOS patients;  
Outcome 01: Ovulation induction; Outcome 02: Pregnancy achievement**



### 2.2.3 Discussion of PCOS Systematic Review

As previously mentioned, properly conducted randomized controlled trials of standard therapy (CC) versus placebo are lacking. Nevertheless, a previously published meta-analyses of existing cross over trials suggested that CC is more effective than placebo for ovulation induction (OR 6.82, 95% CI 3.92, 11.85) and for achievement of pregnancy (OR 3.41 95% CI 4.23, 9.48). However, CC therapy may be associated with twins in 10% of cases, triplets or more in < 1%, rarely ovarian hyperstimulation, a need for in utero insemination to achieve the highest pregnancy rates, and ovarian cancer after prolonged therapy.

The advantages of metformin include a return to spontaneous, monofollicular ovulation induction, therapy avoiding multiple pregnancies and ovarian hyperstimulation. There are no documented risks of ovarian cancer after metformin therapy. Potentially, longer term benefits include cycle regulation thereby avoiding endometrial cancer, reduction of miscarriage rates, and potential risk reduction for type II diabetes. Our results suggest that metformin is more effective than placebo for a return to spontaneous ovulation but not enough data exists with sufficient follow-up time to determine its benefit for achievement of pregnancy. The estimates for metformin versus placebo are, likely, however more reliable than the estimates for CC versus placebo since they are derived from randomized controlled trials. As a comparison, uncontrolled cases series originally suggested that metformin was associated with a better than 90% but a systematic evaluation of properly conducted RCTS suggested the ovulation rate was in the range of 47%.

Our study suggests that metformin may be reasonable first line therapy in otherwise healthy PCOS women who demonstrate oligoovulation with hyperandrogenism and who want to achieve ovulation. For patients who want to achieve pregnancy, however, metformin plus clomiphene citrate may be better than metformin or clomiphene citrate alone. This paper has recently been published in the journal Human Reproduction and is found in Appendix C1.(83)

Recently, Lord and colleagues published meta-analyses with similar results in both the Cochrane Collaboration and the British Medical Journal.(84) The protocol for their paper was identified during our search but the papers were published after the last search date. While their results concerning ovulation and pregnancy were similar to those reported here (6 studies were included in both papers), our paper included 2 different studies and their paper included 5 different studies.(66, 80, 85-89) Two of these five papers were not available to us.(88, 89) Both were requested but unretrievable. One of these appeared to be a cohort study from the abstract.(88) For the other we have no information.(89) Another evaluated coadministration of metformin during recombinant follicle stimulating hormone (rFSH) treatment of patients with clomiphene citrate resistant PCOS.(85) The other two papers had been initially identified but nothing in their title, abstract or methods suggested relevance.(86, 87) Subsequent to Lord's publication, we retrieved these articles which reported mainly changes in metabolic parameters related to metformin. Only 6 of their 13 studies excluded male or tubal factor infertility. The authors included studies where the primary outcome was any of: ovulation induction, achievement of pregnancy or clinical/ biochemical parameters. They include studies which evaluated the latter in their quantitative summary on ovulation and pregnancy rates. Lord did state that they contacted the authors to retrieve missing information. This may have applied to 2 papers and we suspect that one paper was actually included twice secondary to multiple publication.

Nevertheless, our results were very similar to Lord's paper. (Table 5) The fact that two groups independently proposed the question and published the results indicates that the question is of great interest and clinical importance to physicians caring for these patients.

Lord's paper included odds ratios as an effect estimate of metformin versus placebo for both ovulation induction and achievement of pregnancy. Current opinion is that relative risks should be used. In general, effects estimates based on odds ratios are more extreme than those based on relative risks. For comparison purposes, we recalculated our effect estimates using odds ratios.

**Table 5: Comparison of 2 meta-analyses**

	Kashyap et al. 2004	Lord et al. 2003
	RR (relative risk)	OR (odds ratios)
<b><u>Outcome and intervention</u></b>		
<b>Ovulation induction</b>		
Metformin vs. placebo		
Infertile	1.50 [95%CI, 1.13,1.99]	3.88 [95%CI, 2.25,6.69]
Not Infertile	1.45 [95%CI, 1.11,1.90]	
Metformin + C.C. vs C.C.		
Infertile	3.04 [95%CI,1.77, 5.24]	4.41 [95%CI, 2.37, 8.22]
<b>Pregnancy</b>		
Metformin vs. placebo	1.07 [95% CI, 0.20,5.74]	2.76 [95%CI, 0.85,9.98]
Metformin +C.C. vs C.C.	3.65 [95% CI, 1.11, 11.99]	4.40 [95% CI, 1.96, 9.85]

We obtained the following modified results which are closer to Lord's results:

metformin versus placebo for ovulation induction OR 2.83 [1.39, 5.73], metformin + CC versus CC for ovulation induction OR 5.03 {2.35, 10.80}, metformin + CC versus CC for achievement of pregnancy OR 5.05 [1.22, 20.94]. Our results remain incongruent for metformin versus placebo for achievement of pregnancy OR 1.07 [0.17, 6.87]. However, as indicated, different papers were included in the 2 reviews due to our more stringent selection criteria.

Meta-analyses are limited by biases introduced through individual studies as well as biases introduced through the processes of systematic review and quantitative summary. One consistent problem is that none of the studies were powered to assess the most important outcome, pregnancy. The follow-up varied from 1 – 6 months. While oligomenorrhea/ovulation was generally a criterion, studies varied in their assessment and reporting of hyperandrogenemia and clomiphene resistance and this may have produced selection bias. The “cross-over design” used in several studies to assess metformin plus clomiphene citrate versus clomiphene citrate was inappropriate. A properly conducted cross- over trial requires that the treatment can be completely “washed out” without any residual effects after a short period of time; the subjects must

return to baseline with regards to the outcome by the end of washout period so that each subject receives both study treatments and each treatment has an equal chance of affecting outcome. Also, the chance for a successful outcome must not change with time such that the order in which one receives the two treatment options does not affect the rate of success. RCTs also differed significantly in their reporting of methods used to randomize and analyze methods (intention to treat versus per protocol). Compliance, contamination and co-interventions were not reported. Not all studies evaluated infertile PCOS patients and when they did they did not always assess duration of infertility or other contributing factors. We attempted to reduce publication, multiple publication, reporting, conformity and retrieval biases by comprehensive literature searches and searches for grey literature. Only one reviewer completed all steps so extractor and recording bias is possible. We attempted to limit bias by extracting information on three separate occasions.

No randomized controlled trials existed that directly compared metformin to clomiphene citrate for induction of ovulation and/or attainment of pregnancy. This review provides level 1a evidence regarding metformin versus placebo for ovulation induction and pregnancy. The levels of evidence have been defined by the Oxford Center for Evidenced Based Medicine. (Appendix C2) Level 1a included systematic reviews of homogeneous randomized controlled trials while level 1b includes well designed RCTs with narrow confidence intervals. Level 1c consists of all or none case series. Level 2a includes systematic reviews of homogeneous cohort studies, level 2b – well designed cohort studies/ low quality RCTs, level 2c “outcomes research”. Level 4 is similarly divided for case-control studies and level 5 is restricted to expert opinion without explicit critical appraisal. This study suggests that metformin is superior to placebo for ovulation induction in patients with PCOS but that this benefit is not more pronounced with longer therapy (ie > 3 months). The definition of PCOS is very important in determining the group of patients who will have an optimal response. For example, Ng et al. did not find a benefit with metformin. The patients in their study were not overweight ( BMI <23) and also were not hyperandrogenemic as opposed to the other studies where a benefit was found. In a well-defined group of PCOS patients who do not complain of infertility,



metformin also has significant advantage over placebo for resumption of ovulation and regulation of menstrual cycles.

Nevertheless, the data to date do not demonstrate a benefit of metformin versus placebo when the outcome considered is pregnancy. The follow-up time to pregnancy was short and, in the quantitatively summarized studies, pregnancy was not the primary outcome nor were these studies powered to assess pregnancy as an outcome. When we compared metformin plus clomiphene citrate to clomiphene citrate or metformin alone, however, there appeared to be a significant benefit of the combination treatment for both ovulation and pregnancy in patients with PCOS who are both hyperandrogenemic and overweight. However, the comparison of metformin with clomiphene citrate versus placebo with clomiphene citrate was done in each of these studies as a “sequential study” rather than a true randomized controlled trial or cross-over study.

This study was done in preparation for a randomized controlled trial investigating clomiphene citrate versus metformin which was started at Cornell Medical College in August 2002. Subsequently, our study has been terminated but the Reproductive Medicine Network (funded by the National Institutes of Child Health and Human Development (NICHD)) has been conducting a randomized controlled trial investigating metformin, clomiphene citrate, and metformin plus clomiphene citrate. Thirteen centers are involved in the recruitment of 768 patients with oligomenorrhea, elevated testosterone, normal semen parameters and regular intercourse (2-3 times /week) who desire pregnancy. The interventions being assessed include three arms: clomiphene citrate, metformin, and CC plus metformin. This paper demonstrates the need for such a trial. It also provides a basis for the sample size, which was similar in our study when adjusted for two arms.

A need still exists to directly compare metformin and CC as first line agents for ovulation induction and achievement of pregnancy in patients with well-defined PCOS. The currently ongoing study (Pregnancy in Polycystic Ovarian Syndrome – PPCOS) is

designed to answer exactly this question and has the potential to provide valuable insight into the clinical management of this elusive syndrome.

## **2.3 Randomized Controlled Trial of Metformin vs CC in PCOS**

### **2.3.1 Methods for Developing the Randomized Controlled Trial**

The following aspects of the Canadian Institutes of Health Research (CIHR) guidelines were considered in the development of the RCT proposal. (<http://www.cihr-irsc.gc.ca/e/3336.html>)

#### **A. The need for a trial**

Outline the need for the trial and the principle question to be addressed. Summarize current supporting evidence and cite or perform any relevant systematic reviews. Detail the potential clinical impact of such study results and potential use.

#### **B. The proposed trial**

1. Study design - Describe if the trial is experimental or observational. If it is a randomized controlled trial, describe if it is a superiority or equivalence trial.
2. Interventions. - Describe the proposed interventions in both the study and control group in detail. Include the duration of treatment and the duration and frequency of follow-up.
3. Allocation methods - Describe methods used to allocate experimental and standard treatment groups (eg. randomization).
4. Avoidance of Bias - How is bias limited? Are subjects and/or investigators and/or clinicians blinded to treatments? (single vs double blinded) If blinding is not possible indicate why and what other methods are proposed to avoid bias.
5. Inclusion/ exclusion criteria – Identify inclusion and exclusion criteria.
7. Outcome measures – Describe and distinguish primary and secondary outcomes and their measurements. Also describe the measurement of baseline assessments,
8. Sample size – Describe the proposed sample size and the assumptions and calculations used in determining the sample size. Include the minimal clinically important difference and how it was determined. State the power, beta and alpha errors for which the trial is

designed. How are drop outs, contamination and cross-overs handled? Is the sample size based on an intention to treat analysis or per protocol analysis?

9. Recruitment – Describe the planned recruitment rate and preceding investigations or evidence to suggest that this is achievable. Describe how the recruitment is organized, who will be responsible for recruitment and the time period over which it will take place. Describe methods to ensure recruitment on schedule and alternative strategies for recruitment if recruitment is lagging.

10. Compliance, Contamination, Co-intervention and Drop-outs/ Loss to follow-up - Describe anticipated compliance, contamination and co-intervention issues and the evidence for expected numbers of drop-outs or loss to follow-up. Describe strategies to address compliance, contamination and co-intervention issues.

11. Analyses - Provide details of the planned analyses and rationale and adjustments for any additional/ interim analyses. Describe potential subgroup analyses.

12. Economics - If the study will address any health economic issues, describe the proposed analysis and justify the inclusion or exclusion of such a study arm.

13. Timeline - What is the estimated cost and duration of the trial?

### **2.3.1.2 RCT PCOS Protocol**

The methods section of the RCT protocol for the PCOS study was developed as a result of following the CIHR guidelines in section 2.4.2.

#### **2.3.1.2.1 Study Design**

The proposed study is an unblinded, randomized-controlled superiority trial.

#### **2.3.1.2.2 Study Population**

The study population consists of women aged 20 –35 years old with a diagnosis of PCOS by the following criteria: chronic anovulation/oligoovulation (failure to ovulate regularly), infertility, and chronic hyperandrogenism (increased male hormone). Patients will be recruited from the private practices of 9 attending physicians at the Center for Reproductive Medicine, Cornell Medical College.

### **2.3.1.2.3 Inclusion/Exclusion Criteria**

Study participants will be between the ages 20 and 35 at the time they receive the assigned intervention. Oligoovulation is defined as menses less frequently than every 35 days whereas anovulation is defined amenorrhea and inability to document ovulation. Infertility is defined as the failure to conceive for a year or greater. Hyperandrogenism is defined as elevated total and/or free testosterone and/or elevated dehydroepiandrosterone-sulfate. Hyperandrogenism may or may not be accompanied by clinical manifestations. All study participants should have had a male partner with a normal semen analysis (as defined below) and sexual function since the insemination method in this protocol is natural intercourse and not therapeutic insemination.

Exclusion criteria: use of insulin sensitizers or clomiphene citrate within 3 months prior to randomization, abnormal liver or renal function tests or any clinical evidence of hepatic or renal disease, history of diagnosed diabetes mellitus or other medical disease, pregnancy, presence of estrogen dependant tumors, other causes of infertility besides anovulation (i.e. Hypothalamic amenorrhea; male factor, tubal disease, uterine abnormalities or maternal age > 35 years old ), women who smoke or use excessive alcohol, or women with abnormal thyroid function or prolactin tests. Absence of male factor should be documented by a semen analysis which demonstrates normal concentration and motility by Kruger's or World Health Organization (WHO) criteria. Absence of a tubal factor should be documented by normal hysterosalpingogram and/ or laparoscopy with chromopertubation. Investigations for tubal patency, semen analysis, liver and renal function tests, and thyroid function and prolactin tests should be current within 1 year of randomization.

Patients will be recruited from the private practices of 9 attending physicians at the Center for Reproductive Medicine, Cornell Medical College. The primary investigator meets with these physicians on a weekly basis to identify potential eligible participants. Identified patients who met the eligibility criteria were approached by the study nurse. The study nurse and/or primary investigator will explain to potential candidates the nature of the study and, if the inclusion/exclusion criteria are met at the baseline

assessment, obtain informed consent. At New York Presbyterian Hospital, the Institutional Review Board requires that the primary or co –investigators are involved in the informed consent process. Although this is contradictory to principles of clinical trials, it is necessary at this institution. The physician primarily responsible for the care of the patients is not to be responsible for recruitment and therefore we avoid undue coercion. We plan to continue to recruit patients until the sample size was attained (See sample size calculations) or an interim analysis showed a significant difference between treatment arms at which point all subjects, who continue to be in the system, will be switched to the preferential treatment.

#### **2.3.1.2.4 Intervention Protocols**

Progression of treatment through 3 cycles is depicted in Appendix B1. All patients, in both treatment arms, will be instructed to have intercourse 2-3 times per week. Ovulation and pregnancy were determined similarly for treatment arms A and B. Assessment of outcome is detailed below. Patients will be contacted bi-weekly to ensure compliance. All patients were also to receive nutrition counseling. The objective in both treatment arms is for the participant to achieve ovulation and subsequently pregnancy. From the literature, 80% of patients with PCOS achieve ovulation on CC. Forty percent of these patients will achieve pregnancy and 75% of those patients who achieve pregnancy do so within 3 months of an ovulatory dose of CC. Therefore a patient who fails to ovulate at even the maximum dose of CC or who fails to conceive within 3 months of an ovulatory dose of CC is considered a control-treatment failure. The literature suggests that 80 – 95% of patients with PCOS ovulate with metformin within 2 –3 months. We do not have statistics on pregnancy rates but clinical impression of experts as well as physiological intuition suggests that most of these patients who ovulate should conceive also within 3 months if they are going to do so. Therefore failure to ovulate on metformin after 3 months of therapy or failure to conceive with metformin despite 3 months of ovulation will be considered a treatment failure.

The maximum length of time a patient was to spend in Arm A is 5 months. The minimum amount of time a patient was to spend in Arm A is 1 month. If a patient did not ovulate at the prescribed dose in the first cycle, we step up to the next dose of CC in the next consecutive cycle, until ovulation occurs or a maximum dose of 150 mg is reached. Once she achieves an ovulatory dose of CC, we do not increase the dose even if she does not become pregnant. Since 75% of pregnancies occur on CC within 3 months of an ovulatory dose, if 3 months of ovulation with a given dose of CC fail to result in pregnancy, this will be considered a treatment failure.

A patient who enters Arm A starts with CC 50 mg days 5 –9 (post progesterone induced bleed). If this patient ovulates and is pregnant in the first month she is considered a treatment success. If she ovulates but does not become pregnant, she will repeat the CC 50mg course the following month. If she becomes pregnant, this is a treatment success. If she does not become pregnant she will repeat the same course for 1 more consecutive month.

If the patient fails to ovulate the first month at a dose of CC 50mg, the dose is increased to 100 mg in the following month. If the patient ovulates and is pregnant on the first month of CC 100mg, this is considered success. If the patient ovulates but fails to conceive at CC 100 mg, she will repeat CC 100 mg in the next consecutive month. Again if she becomes pregnant, this is considered a success. If not, she will repeat the same protocol the following month for a final cycle. Again, failure to conceive after 3 months of an ovulatory dose of CC represents a treatment (control-treatment) failure.

If a patient fails to ovulate at a dose of CC 100 mg, the dose is increased to 150mg the subsequent month. If she does not ovulate at this dose, she is considered a control-treatment failure. IF she does ovulate and becomes pregnant, she is considered a treatment success. IF she ovulates, but fails to conceive, she continues the same protocol until pregnancy is achieved or a total of 3 months has passed without pregnancy.

Monitoring for ovulation begins at 5, 7, and 10 days post the last CC pill. All patients will be monitored for side effects including nausea, headaches, abdominal pain or bloating, and visual disturbances. Presence of visual disturbances or vascular headaches will require discontinuation of medication.

The maximum amount of time patients will spend in treatment arm B is 6 months. Patients who enter treatment arm B will start metformin at a dose of 500mg daily x 1 week. The dose will be stepped up to 500 mg twice a day the following week, for one week. In the third week the dose will be adjusted to 850 mg twice daily. Stepping up the medications in this fashion is thought to improve compliance. Patients will be monitored for ovulation with blood and ultrasound every 2 weeks until ovulation occurs or the patient has been on metformin for 3 months without ovulation at which point she will be considered a study treatment failure. Once a patient ovulates, she will be tested for pregnancy 14 days later.

Patients will be advised to take the metformin with meals to avoid gastric upset. While on metformin patients will be monitored for side effects including nausea, vomiting, diarrhea and abdominal pain. Patients who develop significant medical illnesses while on metformin will be discontinued from therapy because of the risk of lactic acidosis as mentioned above.

Study participants (Control or experimental) who fail treatment, by the above definitions, will be offered alternative therapy.

#### **2.3.1.2.5 Randomization and Blinding**

Patients will not be allowed enrollment until a diagnosis of PCOS has been confirmed and we have excluded medical conditions that would prohibit their participation. No patients will be randomized until the inclusion/exclusion criteria have been met.

All patients will start medication during the follicular phase. Serum progesterone levels will be used to confirm this and potential pregnancy excluded by use of a serum quantitative hCG drawn at the same time. For patients who have not had a period within

the previous 35 days, menses will be induced with provera 10 mg daily (for 12 days) prior to the start of medication. Once patients are recruited, baseline assessment completed, and inclusion/exclusion criteria met, they will be randomized into one of 2 treatment arms (A or B ) as depicted in the schematic attached in Appendix B1. Randomization will occur in blocks of 6 and randomization will be computer generated. Unfortunately, the nature of the treatments does not allow for blinding of either the patient or the physician. Monitoring of side effects is necessary and also will make obvious the assigned treatment. However, since only hard outcomes are being assessed (ovulation, pregnancy), this should not adversely affect the results. Patients will be assigned a code that will be kept in a central logbook next to their name and their data sheet will be identified, for the purposes of the study, only by the code to protect confidentiality.

#### **2.3.1.2.6 Assessments (baseline and outcome)**

The following demographics will be collected on a data form: code number, age, weight, height, body mass index (BMI), gravidity, term, preterm, aborta, live birth,(GTPAL), medical conditions: Diabetes, renal disease, liver disease; other medications; recent use of insulin sensitizers or oral contraceptive use; duration of infertility; and the following clinical predictors of insulin resistance: waist: hip ratio; waist circumference; acanthosis nigricans; skin tags. Also all patients will have the following baseline bloods drawn (these bloods will be done at the time of routine, requisite blood work to avoid an extra needle stick): total and free testosterone; fasting glucose insulin ratio; LH:FSH ratio; DHEAS; BUN, electrolytes, creatinine, AST, ALT, ALK Phosphatase, TSH and prolactin. These results will also be input on a patient data form. All patients will have a physical exam to rule out co-existing medical conditions. On the day that the patients are randomized they will also have serum progesterone and hCG levels drawn.

Determination of *outcome* will be made the same way in all patients:

1. pregnancy: serum quantitative hCG levels will be drawn on day 28, values > 5 IU/L reflect a positive pregnancy test



2. ovulation: serum progesterone levels will be drawn ; value > 4 ng/ml indicate ovulation

The total hCG assay is an immunoassay using chemoluminometric technology. It is sensitive to 2 mIU/ml. A positive pregnancy results is consistent with > 5 mIU/ml. It is also an immunoassay using chemo-iluminescent technology. The coefficient of variation is 4.2% It is sensitive to a concentration of 0.11 ng/ml. A level > 4 ng /ml is consistent with ovulation. Both assays are regularly calibrated. Outcome assessments will be made as indicated in the schematic Appendix B1. Pregnancy tests will be done 14 days post ovulation, at the time of missed menses, or as clinically indicated. Assessments for ovulation will be made every 2 weeks.

#### **2.3.1.2.7 Compliance/ Contamination/ Co-interventions**

Patients who are recruited will be contacted bi-weekly to ensure they are tolerating the medication. This will allow for early trouble-shooting, although for the reasons previously described, noncompliance is expected to be minimal. Step-wise increments in metformin will also help minimize side effects. Patients will not be able to receive alternate medications without a prescription from their physician. We will monitor for contamination and co-intervention but we really feel these are unlikely to occur.

#### **2.3.1.2.8 Sample size and Recruitment**

##### Sample Size Estimates

The objective of this trial was to show that metformin may be better than clomiphene citrate in achieving pregnancy.

Success rates with clomiphene citrate for achievement of pregnancy in this group of patients are quoted to be about 40%. Our experience at the Center for Reproductive Medicine and Infertility (CRMI) suggests clomid success rates are actually closer to 30%. In consultation with experts we have decided that if treatment rates differ by greater than 15 %, this would be essentially a significant difference clinically. Therefore using the calculation:

$$N = 2p(1-p) (Z_{\alpha} + Z_{\beta})^2 / d^2$$

where  $Z_{\alpha} = 1.968$      $\alpha = 0.05$

$Z_{\beta} = 0.840$      $\beta = 0.20$

$d = 0.15$     is the % of difference below which a difference between treatments is not clinically significant; this is based on the fact that metformin is less invasive and associated with fewer risks i.e. multiple gestation, ovarian hyperstimulation compared to CC

$p = 0.3$

$N =$  sample size per treatment group

A sample size of  $N = 147$  is needed.

Adjustments to the sample size for nonadherence need to be considered.

From the clinical trials in the literature, approximately 5–10% of patients discontinued metformin before planned. However, fertility was not the indication for treatment in these patients. Most of these trials were investigating the effects of metformin on metabolic parameters and menstrual cyclicity. Presumably these patients would have less motivation to continue therapy than infertile patients. There are no reports of CC dropouts in clinical trials for infertility and our experience would support the notion that that would be extremely rare.

Therefore if we adjust the metformin group for an overestimated 5% dropout rate, the metformin  $n = 153$ . If we adjust the CC group for an overestimated 5% dropout rate, the CC treatment group has  $N = 153$ . In summary,  $2N = 306$  will be needed for the study.

#### Recruitment

Patients will be recruited from the Center for Reproductive Medicine, Cornell Medical College. A conservative estimate of the number of patients that could be recruited weekly among the 9 practicing physicians at the Center for Reproductive Medicine is 10 patients per week (accounting for patients who are approached and refuse and patients

who are not suitable). Two of the 9 physicians indicated that they each would see 4-5 new PCOS patients per week. Recruitment will be assessed monthly to investigate whether the objective of a minimum of 40 patients were being recruited each month. With this target, recruitment is expected to be completed in 8 months.

Several of the attending physicians at the primary institution were approached. We discussed whether they agreed on the clinical importance of this trial and whether they would be willing to have their patients participate in such a trial. They are enthusiastic about such a trial especially since different attending physicians have varied opinions about the relative efficacy of these treatments. Also, there are no concurrently ongoing clinical trials that would compete with this trial for recruitment of patients.

To facilitate recruitment, patients will be offered several times during the week, including weekends and evenings, when they can complete the trial prerequisites. The Center itself is open 7 days a week from 0700 am until 1900 pm, and the principle investigator will carry a pager full-time and be available to answer any questions.

If, after 1 month, recruitment is inadequate, alternate sources will be added. The proposal is also to be concurrently submitted to the research ethics committee of the Ottawa General Hospital so that the mechanism is in place to recruit patients at this hospital. A conservative estimate of the number of patients who could be recruited there is 2-4 patients per week. Several attending physicians at the Center for Reproductive Medicine have suggested that if recruitment falls short than an intensive program with flyers and advertisements within the hospital for both patients and attending obgyn staff can be implemented. This strategy has been successful previously.

#### **2.3.1.2.9 Analysis**

An intention to treat analysis will be followed. Although the randomization will be done in blocks, the blocks will not be considered in the statistical analysis.

The following baseline assessment variables will be assessed to ensure comparability of the 2 groups: age, weight, height, BMI, Gravidity, parity, waist: hip ratio; waist circumference, fasting glucose: insulin ratio; total and free testosterone levels.

The outcome variables (ovulation, pregnancy) are dichotomous variables and will be assessed by  $\chi^2$  tests.

Multiple linear regression analysis will be done to determine the relationship between predictors of insulin resistance (BMI, fasting glucose: insulin ratio, and waist-hip circumference) as well as androgen levels (free and total testosterone) and success of metformin versus CC for induction of ovulation as well as pregnancy.

This is a comparative trial. Therefore the data will be analyzed as an intention to treat analysis. “Drop-ins” to therapy will be analyzed according to the treatment to which they were originally randomized. Drop-outs will be analyzed according to the assigned treatment. Patients who become pregnant between treatments will be excluded from the analysis. Patients who become pregnant between treatments will be included in a sensitivity analysis and treated as per the group to which they were assigned. This will help determine if either treatment may indirectly affect outcome even while patients are not actually on the medication.

An interim analysis will be performed when approximately one third of the patients have been recruited. An alpha-spending sequential method (i.e. O’Brien –Fleming) is planned so that the amount of alpha used will depend on the fraction of total information available at the time of analysis. The analysis will be done at time = 1/3 because we expect to have 20% of the total sample size for whom final outcomes are not available concurrently enrolled. The final alpha for tests of significance for the primary outcome at the end of the trial will be adjusted accordingly.

The interim analysis is to serve the following functions:

- 1) to assess for any obvious superiority of one treatment versus the other;

- 2) to allow us to reassess the sample size based on the actual event rate from both groups to ensure the study has adequate power;
- 3) to allow us to check compliance with the protocol;
- 4) to allow us to reassess recruitment.

Indications for termination of the trial include inability to recruit the required sample size or obvious superiority of one treatment over another.

#### **2.3.1.2.10 Data collection and quality control**

Central data collection forms will be used. A duplicate copy is kept in a confidential locked cabinet in the primary investigators office. The primary investigator is responsible for ensuring data collection is up to date and complete on a bi-weekly basis. A study nurse will help to collect the data. The amount of outcome data to be collected is relatively small: ovulation and pregnancy.

As explained above, the center has an excellent mechanism in place for follow-up. Patients come to this center from all over the world for IVF treatment because of an international reputation. Since 1985 only 10 patient outcomes have been lost to follow-up. Because the treatment we are studying is a first line treatment, we expect most of our recruited patients to be from the North-East U.S. Coast.

A training session will be held for all nurses, physicians, lab technicians, and administrative staff who might encounter these patients. Also, monthly luncheons will be held to update these same people about study progress and difficulties.

All assays will be run at the Center for Reproductive Medicine. If for some reason patients need to have blood drawn at a different location, that lab will centrifuge and then freeze the specimen at -20 degrees Celsius and then FED -EX the specimen to us at our expense. This is routinely done for many of our out- of town IVF patients, so a procedure is already in place.

Adverse events will be monitored. Both study drugs are FDA approved. Clomiphene citrate is approved for fertility use. Metformin is not. However, it has already been widely accepted in practice. Available literature does not suggest that these medications are often associated with adverse events or serious side effects. For clomiphene citrate the most serious adverse effects to be monitored are: visual disturbances, vascular headaches and occurrence of ovarian hyperstimulation. Occurrence of multiple pregnancies was also to be monitored. For metformin, common nonserious side effects include nausea, vomiting, and diarrhea in 20% of patients. Both groups will be surveyed via open-ended question if “they have had any unusual symptoms since their last visit.”

#### **2.3.1.2.11 Economic Implications**

No economic evaluations are planned at this stage of the evaluation.

#### **2.3.1.2.12 Avoidance of Bias**

Concealed allocation of randomization will be the strongest mechanism against bias. Treatments are not blinded but outcomes assessed, pregnancy and ovulation, are discrete, hard outcomes measured biochemically so that they are not subject to interpretation.

#### **2.3.1.2.13 Timeline**

The timeline for completion of this trial is 2 years.

### **2.3.2 Results of PCOS RCT**

#### **2.3.2.1 Study implementation**

The study was proposed to the institutional review board in August, 2002. It received approval in October 2002. During the months of June to September 2002, we implemented a strategy to conservatively estimate recruitment. The principal investigator (Sonya Kashyap (S.K.)) tracked new consults each day for each of the 9 participating physicians. Meetings were held with participating physicians, nursing and support staff. A research nurse was recruited and trained to help manage and execute the study. A system was organized where the principal investigator would meet with each of the

individual participating physicians two times a week on consultation days to identify new candidates. Nursing staff was also taught the inclusion criteria so that they may identify candidates which the principal investigator would then screen with the primary physician. The primary physicians were not responsible for recruiting, obtaining informed consent, or explaining the nature of the study to the patients. They were only responsible for the identification of potential candidates.

### **2.3.2.2 Study termination**

By December, despite maximal effort at the initiation of the study, only one patient had been identified for recruitment. This patient declined participation. The study was therefore terminated on the basis of recruitment issues. Issues in study recruitment and design will be reviewed in the discussion section 2.6.2.

### **2.3.3 Discussion of PCOS RCT**

We were very disappointed in recruitment results for this study despite maximal effort. Nevertheless, our analysis of the shortcomings leads us to several lessons for the design of future studies.

A key problem was a change of insurance policies accepted by the hospital during this time. When we had estimated recruitment, we had observed a minimum of 30-40 new, eligible patients per month. We had conservatively estimated that we could recruit 10 of these patients per month. Once the study was started, referrals for this patient population diminished. We later realized that this was likely due to the fact that the university hospital had made a new policy not to accept three insurance programs that they were previously accepting. Since the proposed protocol was low-technology and first line, it is therapy that may be prescribed through either a general practitioners office or a general gynecologists office. With the new change in policy, patients were less likely to pursue a very specialized center where they would have to pay to be treated with low technology therapy that they could receive in their primary care physician's office at no cost to them.

When insurance covered the treatment, more patients presented to the highly specialized center.

Nevertheless, we still were seeing patients that we were failing to recruit. One of our faults was that we overestimated the enthusiasm of the physicians we involved as well as their ability to identify patients. Although the Center for Reproductive Medicine is part of Cornell Medical College, the individual physician functions as in a busy private practice. Some of these physicians may see up to 40 complicated patients a day and could not take the time to identify study patients.

Other problems related to the patient population. Most patients who did come to Cornell did so because they felt that they were in a highly specialized center with an unsurpassed level of expertise. Therefore, many patients preferred not to leave the treatment fate up to a randomized process. Despite the initial enthusiasm for the study, we believe that many physicians also preferred not to leave decisions to the randomization process. We believe this for a few reasons. Despite initial talks that randomization to metformin or clomiphene citrate would be a good study protocol because many of the physicians did not believe the current literature about metformin, this disbelief did not encourage them to enroll patients but rather to still decide on the course of treatment that they preferred. Also, since both therapies are now standard, physicians also felt a pressure to make a recommendation to patients who were referred to them for their expertise and verbalized discomfort with allowing their patients to participate in a randomized trial where both proposed therapies were readily available outside the study. Some physicians later expressed a concern regarding lack of monitoring and in utero insemination for clomiphene patients in the study protocol, despite the fact that the study protocol had been devised with their advice. A lack of control by prescribing a set study protocol as opposed to feeling that decisions were guided clinically was a major downfall.

The Cornell Center for Reproductive Medicine has an international reputation as the leader in the field and therefore receives many international referrals, particularly for highly complicated IVF patients. The centre has very stringent monitoring process for



IVF and several ongoing drug studies for IVF. We had therefore originally targeted a non-IVF population because we felt that there would be less conflict for competing studies. Also, we felt that there was enough controversy surrounding this issue, the standard of care was defined (as opposed to very complicated cases where an expert opinion is needed on an individual basis) and physicians would be more likely to adhere to the study protocol. However, we also underestimated patient enthusiasm. The patients in this protocol are generally young and have an excellent prognosis to get pregnant with either therapy in the long term. Therefore, they were less likely to want to participate in a study where they felt like a “guinea pig”. They generally wanted to be told what would be the best treatment for them. They may have been more willing to participate if they had a poorer prognosis or felt there were fewer viable treatment options available to them. However, a “desperation” factor may also lead to concerns about coercion. Both treatments are now standard of care and are widely available with many primary care physicians. Patients seemed less likely to participate in a study of two standard treatments that they could receive elsewhere. Patients seemed more willing to participate if the study offered an empirical therapy that was not available elsewhere.

Despite difficulties in implementation, the study we undertook was indeed of significant importance. Six months after we discontinued the study, we discovered that a newly found national body called “The Reproductive Medicine Network” in the United States had organized a very similar trial under NIH funding to run as a multicenter trial on the west coast. They had used a very similar protocol and similar MCIDS and sample size calculations. They had estimated a total sample size of approximately 700 patients when they evaluated 3 arms: metformin alone, clomiphene citrate alone, and clomiphene citrate with metformin together. At the time of thesis presentation they had accumulated approximately half of their sample size in 2 years. The fact that such a study is NIH funded addresses the importance of the trial. Challenges that we encountered in a more private setting were overcome by a different style of medical management and financial management on the west coast. Each of the participating centers and principle investigators are known for promoting academic investigation and many have been previously NIH funded for different studies and therefore had previous experience and

documented commitment to and success with randomized controlled trials. Also, west coast medicine tends to be dominated by Health Management Organizations (HMO) and managed care. East coast medicine, especially in Manhattan, is largely private, even at the large academic centers. However, many smaller centers are required to complete such a trial. Cornell would have had the advantage of large numbers at one site. A large IVF program in North America is approximately 300-500 cycles per year. Cornell starts approximately 2200-2400 IVF cycles each year and a similar number of COH –IUI cycles.

Nevertheless, the proposed question is an important one. The current ongoing NIH funded trial by the Reproductive Medicine Network is well-designed and has a good prognosis for answering some very difficult questions about this elusive syndrome and its treatment.

## Chapter 3 Aromatase inhibitors in Poor Responders and IVF

### 3.1 Introduction

During the process of conducting the systematic review on medical induction of ovulation for patients with polycystic ovarian syndrome, a paper was published suggesting that aromatase inhibitors may be of use in this population.(13) This development is even more recent than the development of metformin for ovulation induction. However, upon review of the physiology of aromatase inhibitors, we felt that this category of drugs may be more beneficial in another special group of patients called poor responders. Poor responders are those patients who have a low probability to achieve pregnancy even with in-vitro fertilization. Special characteristics of poor responders may include one or more of the following: age over 40 years; estradiol less than 1000 pg/ml with IVF stimulation; less than 5 follicles/oocytes obtained at retrieval; history of previous elevated estradiol ( $> 75$  pg.ml) or follicle stimulating hormone (FSH) ( $> 12$  mIU/ml DPC assay/  $> 20$  mIU/ml RIA assay); and history of previous cancellation of IVF stimulation for poor response.

We therefore conducted a search for clinical trials (specifically randomized controlled trials) which investigated the effect of letrozole in poor responders both for in-utero insemination and for in-vitro fertilization. We also conducted a search for studies which evaluated the use of aromatase inhibitors in patients with polycystic ovarian syndrome for the purpose of medical induction of ovulation. In the latter group we found several studies but only one randomized controlled trial. (14)We did not find any studies looking at poor responders and IVF but we did find 1 abstract of a cohort study which investigated the use of aromatase inhibitors for the purpose of ovulation induction and one investigating ovulation induction in poor responders.(15, 90)

We had been having difficulty with recruitment for the PCOS study. We discovered this was partly due to a new decision by Cornell Medical College not to accept 3 types of insurance which they had been accepting previously when we did our sample size

calculation. This decision was made one month after the study received institutional review board approval (IRB). This decision significantly reduced not only the total number of ovulation induction cycles we had been seeing but also the number of PCOS patients who were presenting early on for ovulation induction. This institution does not accept insurance for IVF and therefore we have not seen a drop in this patient group. We have been steadily seeing an increase in IVF patients. Also, CRMI has one of the highest success rates for any IVF clinic with poor responders and therefore receives referrals from all over the world. For this reason and the fact that we see at least 200 poor responders every 3 months, we decided to initiate a second randomized controlled trial in this patient population.

Several protocols have been used to try and maximize follicular recruitment in these patients. However, general success rates are still low (<10%). In a natural cycle, the “dominant follicle”, which is the one destined to ovulate, is normally selected by day five. In patients with decreased ovarian reserve, this may occur even earlier, since the responsible hormones rise earlier. Once the dominant follicle is selected, it inhibits the growth of the surrounding cohort of follicles. Attempts to improve follicular response in this group of patients have concentrated on trying to stimulate the ovary to produce more follicles. However, we have not succeeded in delaying the inhibition of surrounding follicles by a dominant follicle nor have we succeeded in delaying the early selection of the dominant follicle in many of these patients. The protocol we suggest here may accomplish this phenomenon.

The primary indication for aromatase inhibitors is for the treatment of advanced breast cancer in women.(91) Aromatase inhibitors prevent the final step in the conversion of androgens to estrogens. Letrozole is a type II aromatase inhibitor. It is a nonsteroidal competitive inhibitor which reversibly binds the aromatase enzyme. By reducing the estradiol output of the follicle, we may prevent inhibition of surrounding follicles and we may decrease the perception of estradiol by the pituitary so that the pituitary increases production of follicle stimulating hormone (FSH) to recruit follicles.

The half-life of letrozole is 45 hours and therefore when it is taken from day 2–7 it should not interfere with the embryo at the time of implantation. Letrozole is metabolized and excreted by both the kidney and the liver. Several authors have investigated the use of letrozole in cycles of ovulation induction for artificial insemination for patients with polycystic ovarian syndrome (PCOS) and also patients with unexplained infertility or previous poor response.(13, 90) Letrozole was found to have a positive impact on the uterine lining when compared to clomiphene citrate in PCOS patients.(14) When used for super-ovulation in normal responders, the number of ampules of gonadotropins and the estradiol levels were decreased without compromising pregnancy rates.(15) Initial reports of pregnancy outcome show no adverse effects of letrozole.(92) For ovulation induction, a 5 mg dose has been shown to be more effective than a 2.5 mg dose without compromising the endometrium.(93)

### **3.1.1 Background**

The first attempts at mammalian in-vitro fertilization (IVF) dated back to the 1800's. In 1930, Pincus documented the first embryo transfer in a rabbit. The first recorded successful IVF was also by Pincus in 1934.(94) For several reasons, this record was considered controversial. The true proof of successful IVF is the live birth of the observation of an in-vitro fertilized pre-embryo. This did not unequivocally occur until 1954 under the direction of Thibault.(95)

The first attempts at human IVF occurred in the 1960's and were initially limited by timing of ovulation, single oocyte retrieval, and low fertilization rates in-vitro. The birth of Louise Brown in July, 1978 marked the first live human birth from a human IVF conceptus and the beginning of an era. The second and third births occurred in Australia and the United States, respectively. The initial two births were the result of laparoscopic, single oocyte retrieval and in-vitro fertilization. Jones et al., however, used human menopausal gonadotropin to stimulate multiple follicular maturation in order to increase the chances of having a viable pre-embryo for transfer.(96)

The evolving progression of utilizing vaginal ultrasound-guided retrieval, controlled ovarian hyperstimulation for multiple follicular recruitment, optimal laboratory conditions, and micromanipulation techniques such as intracytoplasmic sperm injection (ICSI) to overcome male factor infertility have greatly improved the efficiency of assisted reproductive technology. More recently, improved outcomes after embryo cryopreservation and blastocyst culture prevent embryo wastage and hold the promise of reducing the incidence of multiple births secondary to IVF. Preimplantation genetic diagnosis (PGD), as opposed to prenatal genetic diagnosis, allows pre-embryos to be biopsied and evaluated for a multitude of heritable diseases before they are transferred back to the uterus. Oocyte donation allows women with ovarian failure, poor oocyte health, or multiple failed in-vitro fertilization attempts to carry and deliver a pregnancy which carries the genetic contribution of the male partner. Gestational surrogacy offers the potential of women who have major uterine factor (such as congenital absence of the uterus) an opportunity to have their own genetic children.

Utilization of and access to infertility techniques have increased exponentially. The field of reproductive medicine has developed largely around advances in basic science. This was necessary in order to make such techniques efficient. Nevertheless, evaluation of such treatments by properly conducted clinical trials and epidemiological methods is severely lacking. Also, until recently, the utilization of such techniques has been largely unregulated or legislated. Medline reports 499 randomized controlled human trials since 1994. Closer evaluation, however, reveals that most of these are not properly conducted RCTs with appropriate methodology despite the fact that they are reported in this manner.

The Center for Disease Control, in cooperation with the Society for Assisted Reproductive Technology, began collecting information from IVF clinics in the US in 1995. The first complete data collection year was 1996. In that year, 300 clinics reported starting 49 584 fresh, nondonor cycles. In 2001, this number had almost doubled. Three hundred and eighty four clinics reported starting almost 80 000 fresh, non-donor cycles in the US. Between 1996 and 2001 the overall live birth rates per transfer increased 19% for fresh, nondonor cycles and the singleton live birth rate increased 24%

from similar cycles. (<http://www.cdc.gov/reproductivehealth/ART01/download.htm>)  
The majority this increase is due to improvements in success rates for women less than 37 years old.

### **3.1.1.1 Pathophysiology Poor Responders**

Fertility clearly diminishes with age. This has been demonstrated repeatedly. Each female fetus is endowed with a fixed number of oocytes (6-7 million) by 20 weeks gestational age. Through atresia, this number drops exponentially to 300 000 to 400 000 by menarche and 25 000 by age 37-38. Such changes are reflected in the average time to pregnancy. Assisted reproductive technology success rates also diminish drastically with age. Oocyte donation studies illustrate that IVF success depends on the age of the donor rather than the recipient and that oocyte age and health are stronger predictors of outcome than uterine senescence. Donor insemination programs also reflect the fact that female age predicts outcome.

Day 3 FSH and estradiol have been correlated with IVF success in terms of pregnancy rates, number of oocytes retrieved and peak estradiol levels.<sup>(97-99)</sup> For younger patients, FSH and estradiol levels may more accurately reflect a patient's ovarian reserve than her age alone. However, it must be emphasized that normal FSH and estradiol levels in the older patient do not override the impact of chronological age on outcome.

The clomid challenge test (CCCT) can also be used to prognosticate ovarian reserve. The FSH is measured on day three and then again on day 10 following administration of clomid 100mg po on days 5-9. Poor ovarian reserve is defined as a day 10 FSH level greater than 2 standard deviations above the mean. Navot et al. reported that of patients with an exaggerated response only 5.5% subsequently conceived versus 42% of patients with a normal response.<sup>(99)</sup> Some authors have suggested that the CCCT might be more sensitive than basal FSH levels alone but it is not clear whether basal estradiol levels were taken into account in these studies, or whether basal FSH levels were assessed in more than one cycle.

### 3.1.1.2 Epidemiology Poor Responders

Age is the single most important factor that influences IVF outcome. The probability of conception decreases with age, although the fertile window (approximately 6 days before ovulation and 1 day after ovulation) does not change with age. Day specific probability of pregnancy drops 50% for women in their late 20's to their late 30's.(100) This statistic does not include the increased spontaneous abortion rate associated with sporadic genetic abnormalities associated with conceptuses of older patients. Age is the strongest predictor of ovarian reserve. Reproductive success diminishes exponentially with advancing maternal age. The classic studies of the Hutterite population evaluate a society where the average age of marriage is 22 and the only fertility barriers include lactational amenorrhea and a natural age and parity related decrease in coitus.(101) Donor insemination programs also demonstrate a female dependant decrease in fecundity. A recent epidemiological study of almost 800 couples practicing natural family planning revealed that female fecundability begins to decline in the late 20's and male fertility demonstrated a less, although significant decline in the late 30's. Biochemical measures of ovarian reserve are available which may predict a patient's prognosis and response to stimulation. Day 3 FSH and estradiol have been correlated with IVF success in terms of pregnancy rates, number of oocytes retrieved and peak estradiol levels.(97-99) For younger patients, FSH and estradiol levels may more accurately reflect a patient's ovarian reserve than her age alone. However, it must be emphasized that normal FSH and estradiol levels in the older patient do not override the impact of chronological age on outcome.

In 1996, 8412 fresh, nondonor cycles were started in women over the age of 39. While "diminished ovarian reserve" was not an established, reported diagnostic category in 1996, 9% of women treated with assisted reproductive technology had a singular diagnosis of diminished ovarian reserve in 2001. This statistic is likely a very conservative estimate since 27% of cycles were started in women over the age of 39 in the year 2001 versus 17% of cycles started in women over the age of 38 in 1996. [8412 cycles started in women over the age of 39 in 1996, 23327 cycles started in women over



the age of 38 in 2001] With increasing patient education, success rates, and accessibility to infertility services the utilization of such services has been rapidly increasing. Also, with increased stresses, education and career demands, many women and couples are delaying child bearing and this contributes to the larger proportion of women presenting for ART for decreased ovarian reserve.

### **3.1.1.3 Clinical Treatment Options, Success Rates and Evidence**

Women with decreased ovarian reserve have few options. Controlled ovarian hyperstimulation and intra-uterine insemination are not an effective treatment in this patient group. Success with in-vitro fertilization is also very limited in this group. At each step of the process, results are modest. Response to ovarian stimulation is limited, numbers of eggs obtained at retrieval are often few, fertilization may be compromised as a result of aged oocyte health, the number of embryos available for transfer is usually small and embryo quality is an unknown factor. Early pregnancy miscarriage rates are also high. Perhaps the most difficult step is poor response to ovarian stimulation.

Despite poor prognoses and lack of rigorously evaluated stimulation protocols in this patient group and associated high direct costs to patients, utilization of such services, as described above, illustrates the extremely visceral desire to procreate ones own genetic offspring. Other alternatives include natural cycle IVF, with extremely low success rates, in patients who fail to respond to stimulation, donor egg, adoption, and the alternative to do nothing.

Pregnancy rates per fresh, nondonor cycle started in women older than 39 years for the year 1996 were 13.4% with a live birth rate per cycle of only 8.7%. The cancellation rate in this group was 22.3% versus 9.4% and 15.3 % in women aged < 35 and 35-39 respectively. In 2001, 14% of all cycles (11349) were cancelled and 84% of these (9533) were cancelled for poor response. In that year, success rates were reported for women aged 38 –40, and 41-42. For women more than 38 years old, the percentage of cycles resulting in pregnancies was 43.5% and the live birth rate was 30%. The percentage of cycle cancellations in this group was 39.2 versus 9.6 and 14.1 in women < 35 and 35-37 respectively. Donor egg offers couples with poor ovarian response, ovarian failure, or

genetic diseases the option of gestating a pregnancy with the paternal genetic contribution of the couple. In 2001, 7772 fresh donor egg transfers were completed in women of all age groups. In 1996, 3822 fresh donor egg transfers were completed in women of all age groups and 2472 of these were in women over age 39. The success rate of such transfers is dependant on the age of the donor rather than the recipient and is therefore generally quite high. For example, in 2001, 47% of fresh transferred embryos (from donor eggs) resulted in a live birth (much higher percentage may have actually resulted in clinical pregnancies). Nevertheless, utilization of such services is limited by many factors not the least of which are accessibility to a suitable donor and personal attitudes towards acceptance of donor gametes. In 2001, 89% of reporting clinics offered donor egg services. The desire to bear one's own genetic progeny, however, must be extremely visceral when one considers the differences in success rates between donor egg cycles and nondonor cycles in this patient group, the relatively high direct patient financial, physical and emotional costs versus the fact that most of these patients still prefer to do nondonor egg cycles.

A medline search for randomized controlled trials evaluating IVF protocols in poor responders identified 16 papers. Of these 16 papers, 4 reported results on stimulation protocols for IVF, 3 evaluated the addition of growth hormone on steroid parameters and/or IVF outcome.(102-109) One evaluated the role of assisted hatching in women of advanced maternal age but this was actually a cohort study, not a randomized controlled trial.(110) The other papers were either letters, or studies of non-poor responder patients. The "randomized controlled trials" that do exist suffer from similar problems. The largest study is 48 patients and it acknowledges a need for larger studies. All of these studies failed to define a primary outcome (ie follicle number, fertilization rate, or pregnancy?), sample size, and minimally clinically important difference. Also, the methods of "randomization" are not defined, ie stratified, blocks, computer generated, or concealed, and often appear to be quasy randomized (eg serial recruitment, independent physician decision, health code or clinic day). Most of these studies fail to show a significant difference between protocols but in fact do not have sufficient power to make conclusions. In addition, the definitions of "poor responders" are lacking and

inconsistent. An effort to standardize this definition is required in order to compare data from different institutions. The volume of “poor responder” patients is often limited at any one institute. At the CRMI, 14% of patients were reported to the CDC to have a singular diagnosis of diminished ovarian reserve. However, more than 700 cycles were started (> 30%) in patients over the age of 38.

Nevertheless, considering the physical, financial, and emotional burden associated with the treatment of such patients, appropriate randomized controlled trials are necessary to prove the efficiency of therapy. Such trials are difficult and often impossible to implement. In-vitro fertilization is a completely uninsured service in 37 states, partly insured in 5 states, and completely insured in 3 states. The direct costs to the patients are high, averaging almost \$10000 in the US in 2001.(2) In addition, patients have often undergone multiple cycles of less invasive treatment before being referred to an IVF centre and therefore many of these patients are very anxious. Many patients have undergone IVF therapy in other centres before moving to a center with a particular expertise in this particular patient group. The capitalistic nature of competing business interests for IVF clinics, particularly in densely saturated areas, also discourage groups from organizing or participating in such trials. Waiting times, regression to the mean, and alternative treatment plans also interfere with the ability to conduct such studies.

Current standard protocols for IVF in patients who are considered “poor responders” include flare and antagonist protocols. Nevertheless, no properly conducted trials exist to define the benefit of these protocols in general or compared to one another. Our study compares a newer, experimental protocol to these already standard protocols. It will be the first study to rigorously evaluate a new protocol for poor responders against existing protocols.

#### **3.1.1.4 Rationale for a Feasibility Study**

One of our original objectives was to propose an important question in reproductive medicine that could best be answered by an RCT. Our first question, although important, was not successful when we implemented our RCT. Nevertheless, that same question is

snow being addressed in a multicentre RCT conducted in the US by the Reproductive Medicine Network and funded by the NIH. We learned many lessons from that attempt at a randomized controlled trial. One of these was that an RCT in our specific population may not go over well unless one of the therapies is an experimental therapy not otherwise available to our patients. An RCT comparing two accepted, standard therapies was not well-accepted in this population. Also, patients who have excellent prognoses to achieve pregnancy regardless of therapy may be less likely to participate. The poor responders constitute a group of patients who have few therapeutic options and therefore may be more receptive of experimental protocols. In addition, many unforeseen pitfalls may occur while conducting the study. Some of these include overestimated physician enthusiasm for the study, lack of anticipation for factors which may significantly affect patient recruitment, and failure of participating physicians to adhere to study guidelines partly due to inflexibility of clinical management study protocols. Therefore, instead of attempting to conduct a full study, a feasibility trial may be a more appropriate first step.

Given financial and emotional costs incurred by this group of patients called “poor responders”, we felt a properly conducted RCT to evaluate a new, although quickly disseminating therapy, was long overdue. The use of aromatase inhibitors in this group has not yet become standard of care but is perhaps the newest stimulation protocol to be proposed. We therefore conducted a systematic review of the subject before designing and implementing a feasibility trial. The protocol that follows outlines the design and sample size for the whole trial based on an expected clinical pregnancy rate of 10% and an minimal clinically important difference (MCID) of 30%. Given the lessons learned from the previous attempt at an RCT, we decided to first implement a feasibility study. The objectives of the feasibility study would be:

- 1) To ensure that patients can be recruited. Two issues arise. One issue is that the type of patients (considering inclusion/exclusion criteria) can be recruited. Second, that the number required for the study can be recruited. Previous experience suggests that both unanticipated patient and physician attitudes may influence recruitment.

- 2) To determine if we are able to randomize patients given the patient population and environment. Since the treatments are not blind, it is essential that the allocation is concealed. We suspect there will be some pressure to direct allocation.
- 3) To estimate the expected pregnancy rate in the group of patients being recruited for the study. We expect a 10% pregnancy rate in this group of poor responders presenting to our centre for specialized care. In fact, in women over the age of 40 without a previous history of stimulation or high FSH, our centre has seen up to 30% pregnancy rate. The target group of patients for this study however is much more clinically challenging .
- 4) The final objective of this feasibility study is to run a pilot version of the study to identify other logistic problems before a full protocol is designed and implemented.

### **3.1.2 Objectives**

#### **1) Systematic review:**

- a) To perform a systematic review using, Cochrane, MOOSE and Quorum guidelines (RCTS and cohort studies), and provide a quantitative summary where applicable, regarding the following question:

Question: Are aromatase inhibitors, in combination with gonadotropins, beneficial compared to standard, aggressive protocols for IVF treatment of poor responders?

#### **2) Randomized controlled trial:**

- a) To design an RCT protocol and implement a feasibility study on the same question, including the above objectives for the feasibility study.

Question: Are aromatase inhibitors, in combination with gonadotropins, beneficial compared to standard, aggressive protocols for response to ovarian stimulation and improvement of pregnancy outcomes in IVF for poor responders?

## **3.2 Systematic Review for Aromatase Inhibitors in Poor Responders**

### **3.2.1 Methods for conducting a systematic review**

#### **3.2.1.1 Search strategy**

A **computerized literature** search of the following databases, using an OVID vendor and Polaris interface, was done: Medline, Premedline, Cochrane Register of Clinical Trials, and EMBASE. The databases were searched for the last 25 years. MeSH heading and textwords, for databases without Mesh headings, were used. Adjacency operators, and truncation were employed. A preliminary search had already been conducted to maximize potential keywords. We did not apply a study filter and the search will not be limited by language or year of publication. It was repeated at two-week intervals until May, 2004.

A **hand search** of ten years of the following journals was also completed: Fertility and Sterility, Human Reproduction, Journal of Clinical Endocrinology, New England Journal of Medicine, the Lancet, and the American Journal of Epidemiology. Online indexing facilitated this process.

To avoid publication bias, we searched for **grey/unpublished literature**. Three content experts were contacted. Also, 10 years of conference proceedings for the following were searched: The Canadian Fertility and Andrology Society, The American Society for Reproductive Medicine, and the Society for Gynecologic Investigation.

The Cochrane database was searched for systematic reviews, study protocols, or completed, registered clinical trials. Titles and abstracts were screened and articles retrieved if they passed the relevance filter or if there was uncertainty whether or not they were relevant. Bibliographies of review articles, systematic reviews, and retrieved articles will also be searched for candidate articles. Retrieved articles were reviewed for inclusion/exclusion criteria and those articles meeting these criteria were kept for critical appraisal and data collection.

**Table 6: Poor responder search strategy**

<b><u>A. Key Words</u></b>	
<b>Study population</b>	<b>Intervention</b>
Poor adj responder\$	Aromatase adj inhibito\$
IVF	Anastrozole
In-vitro adj fertiliza\$	Letrozole
In adj vitro adj fertiliz\$	Testolactone
In-vitro adj fertilis\$	
In adj vitro fertilisa\$	
ART	
Assisted adj reproduc\$	
Ovar\$ adj stimula\$	
<b><u>B. Databases</u></b>	<b><u>C. hand search of journals and conference proceedings*</u></b>
Cochrane controlled trials register	Fertility and Sterility
Cochrane database of systematic reviews	Human Reproduction
EMBASE ( 1980 – 2002)	New England Journal of Medicine
Premedline	Journal of Clinical Endocrinology and Metabolism
Medline (10 years)	Lancet
Current Contents	Journal of the American Medical Association
Biological abstracts	Cochrane Database
	American Society of Reproductive Medicine
	*facilitated by online indexing

### 3.2.1.2 Inclusion/ Exclusion

One reviewer has completed all stages. MOOSE and QUOROM guidelines were followed.(6, 7) Titles and abstracts were screened and articles retrieved if they passed the relevance filter or if there was uncertainty as to whether or not they were relevant. Bibliographies of review articles, systematic reviews, and retrieved studies were also searched for candidate articles.(71) Retrieved articles were then reviewed for inclusion/exclusion criteria. Those articles that met the criteria were then kept for critical appraisal and data collection. The reviewer was not blinded at any point to the authors or sources of publication as the evidence for such blinding is weak and the reviewer was likely to be previously familiar with some of the literature. (72)

### 3.2.1.3 Quality Assessment

The Jadad scale was to be used for quality assessment of randomized-controlled trials.(73) Newcastle-Ottawa scales for quality assessment were to be used for cohort studies. Quality assessment involves evaluation of patient selection, assessment of exposure or outcomes, administration of interventions, and controls for confounding factors. (Appendix A)

### 3.2.1.4 Study Inclusion (PICOS)

“**PICOS**” defines the inclusion criteria for a systematic review. The acronym represents: types of **P**articipants/**P**opulation, types of **I**ntervention, types of **C**omparison, **O**utcomes and types of **S**tudies.

**Study population:** The Study population of interest includes women of reproductive age (age 20 to 45) who have a history of infertility due to decreased ovarian reserve (as defined by previous poor response to stimulation, age, previous increased FSH or age > 40).

**Intervention:** The interventional drugs of interest are aromatase inhibitors alone or in combination with gonadotropins for IVF compared to standard, aggressive protocols.



**Comparisons:** The comparisons of interest include the following: aromatase inhibitors plus gonadotropins for IVF versus coflare or antagonist IVF ovarian stimulation protocols.

**Outcome:** The primary outcome is pregnancy rate; secondary outcomes include oocyte numbers, number of embryos available for transfer, cancellation rates, estradiol levels at day of hCG, fertilization rates, and endometrial thickness.

**Study design:** Both randomized –controlled trials and cohort studies were evaluated. A quantitative summary was to be done where appropriate for homogeneous studies but RCTS were to be evaluated separately from cohort studies. Different treatment options were to be compared separately.

### **3.2.1.5 Quantitative Methods**

Where appropriate Revman 4.1 and Metaview 4.0 were to be used to analyze data. However, the data available was extremely limited for this novel therapy. Therefore, the results are presented descriptively and there is no quantitative summary.

## **3.2.2 Results of the systematic review**

### **3.2.2.1 Study inclusion**

Although we did not limit the language of publication, all retrieved articles were in English.(111, 112)

All initially retrieved articles and abstracts were published between 2001 and 2003. An updated search in June, 2004 revealed 1 additional abstract.(113) ((13, 14, 114)

None of the identified studies qualified for quantitative or qualitative systematic review. Most studies were uncontrolled case series which evaluated aromatase inhibitors in women with PCOS who were undergoing COH-IUI. Two randomized controlled trials evaluated aromatase inhibitors in controlled ovarian hyperstimulation and in-utero insemination. No RCTS evaluated the use of aromatase inhibitors in poor responders either for COH –IUI or for IVF. One case series evaluated aromatase inhibitors in poor responders undergoing COH-IUI. The recent abstract of aromatase inhibitors in IVF

patients with poor response to ovarian stimulation was a case series and observed a 29% pregnancy rate versus 9% in the same patients in their previous cycles.

### **3.2.2.2 Characteristics of Included studies**

No studies evaluated the use of aromatase inhibitors in IVF or poor responders. No studies qualified for inclusion in a systematic review.

### **3.2.2.3 Meta-analysis**

No quantitative summary is possible.

## **3.2.3 Discussion of the Systematic Review**

No randomized controlled trial or cohort studies exist comparing the use of aromatase inhibitors in patients with poor response to ovarian stimulation who are undergoing IVF to standard aggressive protocols for IVF such as the coflare and antagonist protocol. The paucity of available information may be related to two issues. First, the proposed protocol is novel. Aromatase inhibitors were initially suggested for ovulation induction for women with PCOS 3 years ago. To our knowledge, this was the first proposed trial of aromatase inhibitors for poor responders in IVF. Second, the burden of disease associated with infertility and a low response to ovarian stimulation is high. These patients have often undergone multiple failed attempts before reaching a center with an expertise in this area. The stimulation is often tenuous and requires increased frequency of monitoring. For subjects who are likely to produce less than 5 oocytes, many physicians and patients are reluctant to experiment with protocols, especially in a randomized fashion. Additionally, for patients who have failed to respond to previous protocols and are undergoing one last attempt, both physicians and patients may opt to use an experimental protocol directly rather than faced a randomized process.

Nevertheless, we feel that given the fiscal and emotional burden of disease, such a trial should be undertaken before aromatase inhibitors are widely implemented. The purpose of this systematic review was to prepare for such a study.

### **3.3 Poor Responder RCT**

The same CIHR guidelines were followed as in section 2.4.2. A full protocol is found in the results section 3.5.2.

#### **3.3.1 RCT protocol for Aromatase Inhibitors in Poor Responders**

The following methods section of the RCT protocol for Poor Responders was developed as a result of the CIHR guidelines for randomized controlled trials.

##### **3.3.1.1 Study Design**

We will conduct an open randomized controlled trial in which patients will be randomized to the study protocol versus a standard protocol chosen by their primary physician. Only patients who have already scheduled an IVF cycle will be considered for study. The patients and physicians will not be blinded to the treatment. The outcome to be assessed is a “hard outcome” - ie. pregnancy and the study allocation will be concealed.

##### **3.3.1.2 Study Population**

Infertile women between the ages of 20 and 45 who have scheduled an IVF cycle at the center for reproductive medicine and who have been prescribed one of two aggressive protocols for poor responders (flare or antagonist protocol). Patients will be recruited from the practices of 9 physicians at the Cornell Institute for reproductive medicine.

##### **3.3.1.3 Inclusion/Exclusion Criteria**

Poor responders will be chosen on the basis of having already been prescribed one of two aggressive poor responder protocols and one or more of the following inclusion criteria:

- 1) history of previous stimulation for IVF resulting in fewer than 5 eggs retrieved
- 2) estradiol pre-administration of human chorionic gonadotropin of < 1000 pg/ml in a previous IVF cycle
- 3) history of cancellation before retrieval in a previous IVF cycle for poor response
- 4) age greater than 40 years old
- 5) previous high day 3 follicle stimulating hormone (FSH) or estradiol (E2)

Each of the above inclusion criteria identifies a subset of patients who are most likely to have a suboptimal response to stimulation. This is the group of patients we would like to target. Criteria 1 –3 identify women who have previously had a poor response to stimulation and failed IVF treatment. Criteria 4 and 5 predict a poor response and this has been well documented in the literature. As mentioned earlier, advances in ovarian stimulation for poor responders have been extremely limited. The theoretical risks of the described protocol are very limited in this group of patients especially since the half life of letrozole means it will be completely out of the system by the time of retrieval, much less transfer. Although it is an experimental protocol, there are several published and ongoing studies and it is quickly becoming an accepted standard treatment.

Exclusion criteria: Patients with a history of the following will not be considered:

- 1) patients with a history of failed fertilization with intracytoplasmic sperm injection (ICSI) in previous IVF cycles
- 2) patients with a history of implantation failure (endometrial/uterine defects) as the primary cause of IVF failure
- 3) patients with a history of recurrent miscarriage
- 4) patients with a history of lipid disorders
- 5) patients who are currently pregnant
- 6) patients with a history of abnormal liver or kidney function
- 7) couples who plan to undergo preimplantation genetic diagnosis
- 8) couples who are planning to undergo co-culture for the first time
- 9) couples who are undergoing TESE (testicular epididymal sperm extraction) without donor backup (successful TESE with ICSI will result in 70% fertilization rate; however, patients who do not accept donor back up in the event that no sperm is retrieved, do not have any chance at fertilization and embryo transfer in that case).
- 10) poor responders who have previously conceived with a given protocol and who therefore are to be placed only on that protocol again.

### **3.3.1.4 Intervention Protocols**

We will obtain consent prior to randomization or any procedures. Consent will be obtained in person by one of the principle investigators in accordance with the rules of the local institutional review board. The patients' primary physician is not be responsible to get the consent in order to avoid coercion.

Patients will be randomized to treatment arm A or B (Figure 9) prior to the day on which stimulation is scheduled to start. This is necessary in order for patients to receive their medications in time to start the cycle. The cost of medications is not be more than in a standard IVF cycle and for those patients randomized to the letrozole arm, the cost may be less than with the standard protocols since letrozole is relatively inexpensive (costs vary by pharmacy but estimated at approximately \$10 per pill the total cost for the letrozole will be approximately \$50) compared to GnRH antagonists and agonists which would be used in the standard protocols. The costs of the medications and procedures are the financial responsibility of the patients as per usual for any IVF cycle. Subjects are to spend a maximum of 2.5 months in either treatment arm, from the time of randomization to detection of an intrauterine pregnancy by ultrasound.

If a subject agrees to participate and meets the inclusion/exclusion criteria, blood work (Follicle stimulating hormone, luteinizing hormone, and estradiol levels- approximately 1 tablespoon) will be done on day 2 of the menstrual cycle. If these levels are not elevated, she will proceed.

Patients in group A will take letrozole 5 mg orally day 2 – 6 of their menstrual cycle. (Figure 9) On day 4 of stimulation, we will overlap gonadotropins (FSH (Follistim® or Gonal –F®) plus human menopausal gonadotropin (hMG -Pergonal® or Repronex®)). The dose of gonadotropins will be adjusted daily on the basis of bloodwork and ultrasound, as per standard procedure. No excess blood tests or ultrasounds are attributable to the study protocol versus the standard protocols. Patients routinely have blood tests drawn every 1 –2 days during stimulation and then they have a pregnancy test on approximately day 28. Usually in a standard IVF cycle the average number of blood

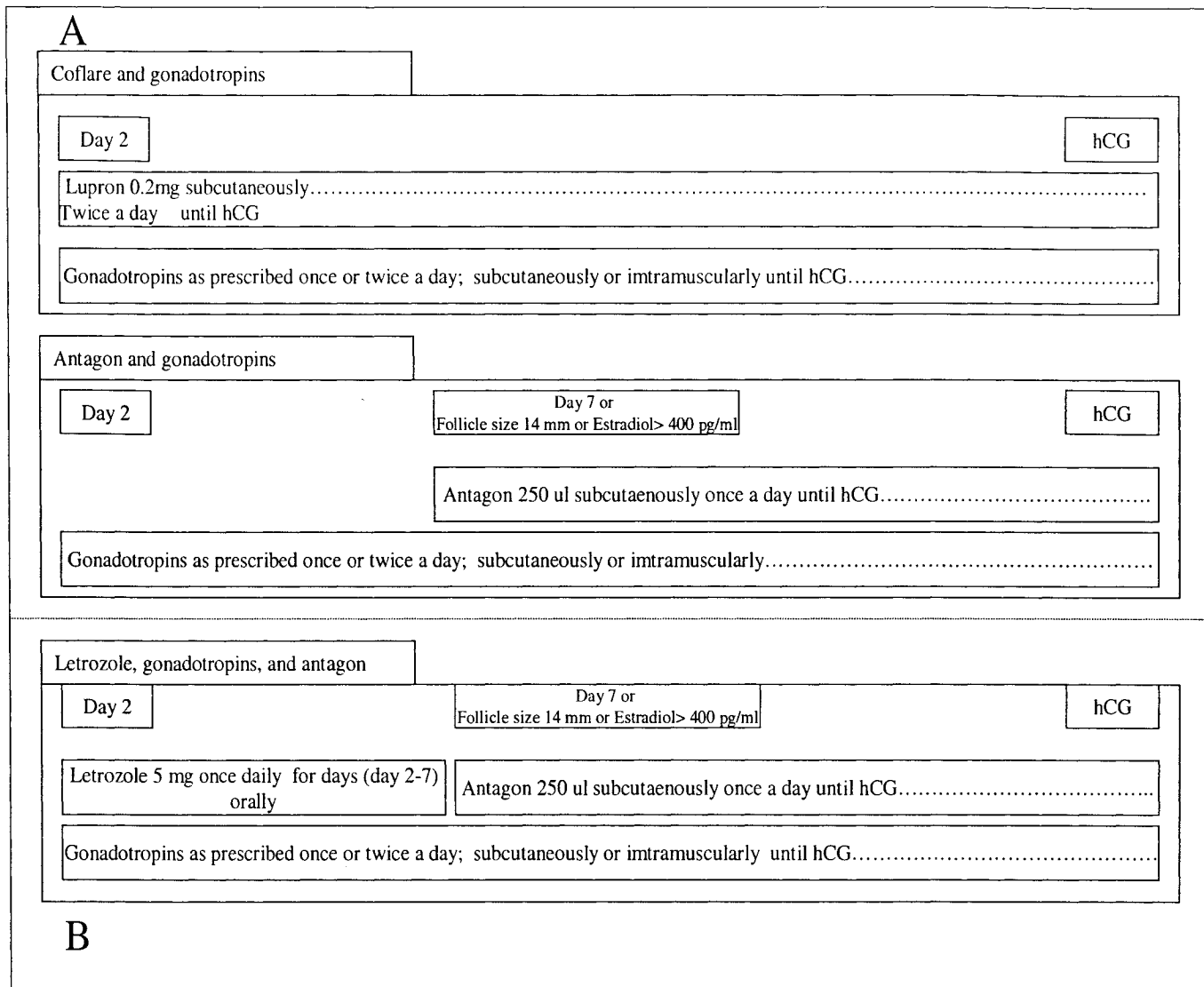
tests is 12 and it will not be different as a result of the study. The number of ultrasounds will be approximately 10, as clinically indicated and again it will not be different due to the study. Patients will then receive human chorionic gonadotropin (hCG) when their follicles reached 16 –18 mm transvaginal measurement, as is standard. Surgical retrieval will occur 34-36 hours later and oocytes (eggs) will be fertilized according to the method which is clinically indicated by the patient's history (ie. Intracytoplasmic sperm injection (ICSI) versus routine insemination). Patients will return for embryo transfer on day 3 post retrieval.

Patients who are randomized to protocol B will follow the medication protocol otherwise prescribed by their physician. This is one of 2 standard protocols: GnRH antagonist with gonadotropins or coflare/ microflare GnRH agonist with gonadotropins. They will also start on day 2 of their menstrual cycle and medications will be adjusted as is clinically indicated by their estradiol levels and transvaginal ultrasounds. They then will also receive hCG at follicular sizes of 16-18 mm for retrieval 34-36 hours later, and the method of fertilization will be determined by clinical history. They will return for a day 3 embryo transfer.

The only difference between the 2 protocols will be the addition of letrozole in the study protocol from day 2-6. The number of ultrasounds and the frequency of bloodwork will be the same. The risks to the patients are those of the risks of IVF in general: the risk of infection or bleeding secondary to the oocyte retrieval (0.3 – 3 %); the risk of injury to abdominal organs or blood vessels during the procedure (< 1%); and the risk of anesthetic for the retrieval procedure (< 0.1%). These risks are not increased as a result of the study since all of the subjects have already planned to do IVF. The risk of ovarian hyperstimulation and or multiple pregnancy (twins or triplets) is extremely low (<< 2%) in this specific patient group whose pregnancy rate is estimated to be 10%. Again, this is the risk associated with IVF and should not be increased by the study. The use of letrozole can be associated with nausea and vomiting (11%). Longer term effects of aromatase inhibitors such as decreased bone density and adverse effects in the lipid profile have only been seen with long term treatment (> than 60 days). Our patients will

only take the medication for a total of 5 days. Other uncommon side-effects from prolonged use of Letrozole include: headaches (13%), fatigue (11%), muscular aches, hot flushes (6%), and hair thinning. All of these effects are uncommon and are associated with longer use of letrozole than in this study. Also, all side effects normalize with discontinuation of medication.(91) Letrozole is known to be teratogenic during pregnancy in rodents. However, the medication is only to be taken for 5 days during stimulation and should be out of the patients' system long before the time of hCG, let alone embryo transfer. Since all patients will have day 2 Estradiol, FSH and LH levels plus an ultrasound done prior to stimulation, any spontaneous pregnancies will be diagnosed prior to the patient starting the medication and therefore exposure avoided.

**Figure 9: Poor responder ovarian stimulation protocols for IVF**



**3.3.1.5 Randomization and Blinding**

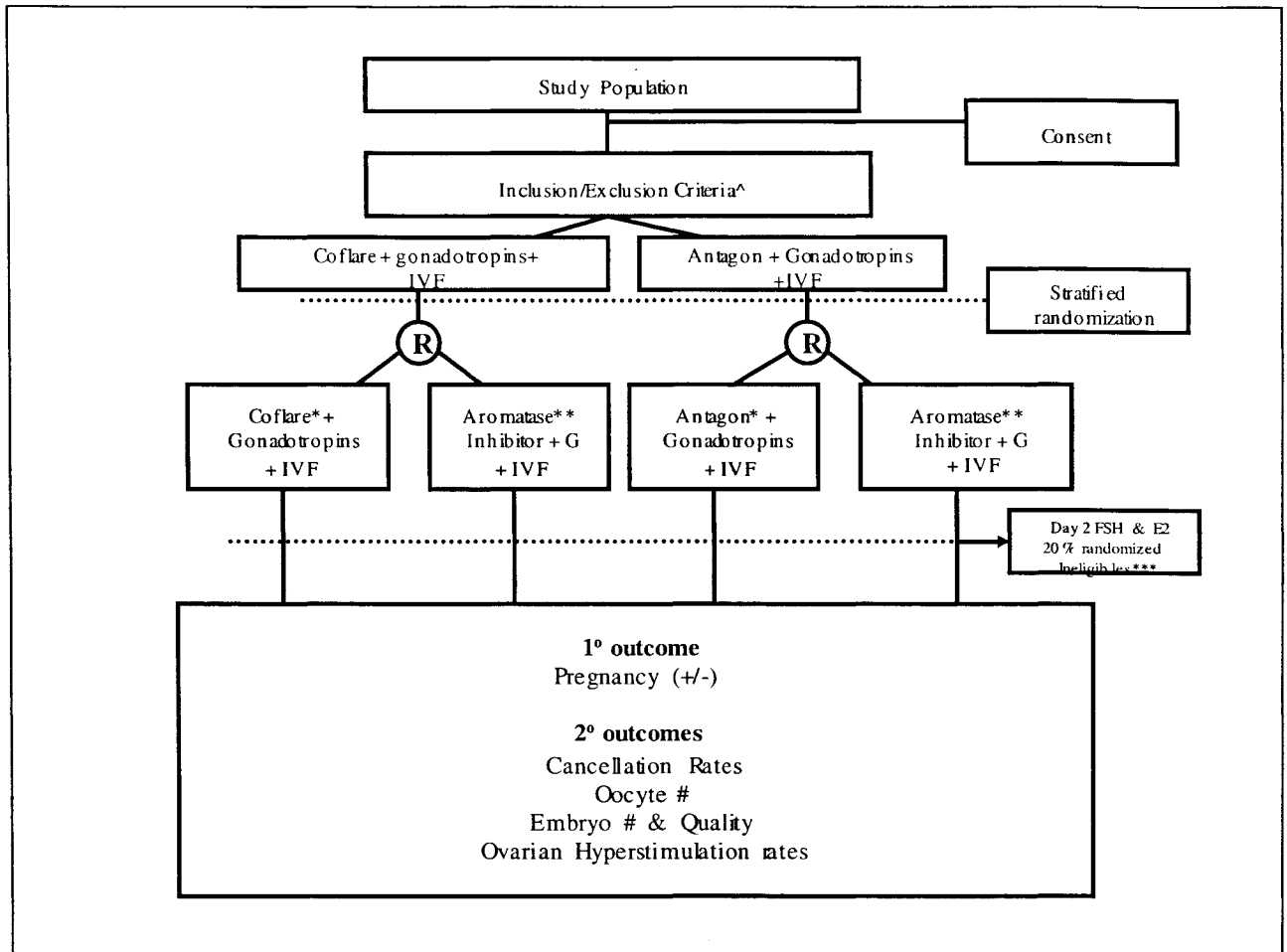
Randomization schedules will be computer-generated and use blocks of 4 and 6. Randomization envelopes will be created by an outside independent group, not associated with the study, so as to avoid bias by the principal investigator. Patients will be recruited on the basis of: 1) assignment to an aggressive stimulation protocol (either antagonist or coflare) and 2) the above enumerated inclusion criteria. Since patients are assigned to one of 2 protocols prior to recruitment and randomization, we stratified the randomization so that patients will be placed in one of 2 strata based on their previous



protocol assignment and then blocked randomization in each of these pathways. (Figure 10) Unfortunately, the nature of the treatments does not allow for blinding. The coflare/antagon protocol require injectable medications and the letrozole protocol requires fewer injectable medications but also requires oral medications. Although blinding is not possible, hard outcomes (pregnancy rates, number of oocytes retrieved and fertilization rates) and extra precautions will be taken to ensure concealment of allocation and compliance, blinding is not necessary.

Patients will be identified according the above inclusion/ exclusion criteria. However we expect approximately 20% of randomize patients would be deemed ineligible by their day 2 blood tests. This is, unfortunately, unavoidable since logistically it is essentially impossible to wait and randomize patients after the day 2 blood test since the results are not obtainable until approximately 1600 in the afternoon. Patients must start their medications on the same day in order to maximize their chances of optimal stimulation. Again, the criteria for not starting the cycle are hard, nonflexible criteria. Patients who have elevated FSH ( $> 20$  mIU/ml by the radioimmunoassay and/or  $> 12$  MIU/ml by the DPC assay- we run both on all patients) will not be allowed to start. Neither patients, nor the physicians can override these criteria. The 20 % randomized ineligible figure has been incorporated into the sample size.

**Figure 10: Poor responders RCT protocol**



**3.3.1.6 Assessments (Baseline and Outcome)**

Baseline assessments include age, BMI, number of previous failed IVF cycles, history and level of high FSH (follicle stimulating hormone – mIU/ml) on day 2 and /or 3, number of previous high FSH levels, history and level of elevated estradiol (pg/ml) on day 2 and /or 3, baseline day 2 blood tests for study cycle done at CRMI ( FSH, Estradiol, LH) and baseline day 2 ultrasound done at CMRI for endometrial stripe thickness and ovarian morphology.

Outcomes to be measured are as follows:

Primary outcome: Pregnancy is determined by the presence of a gestational sac in ultrasound which will be performed between week 5 and 7 as defined by the Society for Assisted Reproductive Technology.

Secondary outcomes will include peak estrogen levels, doses of gonadotropins required, number of oocytes retrieved, embryo number and quality, endometrial stripe thickness on the day post human chorionic gonadotropin administration (hCG).

Maternal age and history of previous failed cycles will be investigated as possible confounders.

### **3.3.1.7 Compliance/ Contamination/ Cointerventions**

Patients who are undergoing IVF are extremely motivated and are required to comply with a stringent monitoring protocol. These patients are assessed with blood tests and ultrasound every 1 –2 days as is clinically indicated. In addition, IVF is very costly and these poor responders know that they have exponentially declining success rates with time and therefore are unlikely to drop out. Compliance is measured by repeated visits (every 1 -2 days), daily nursing phone calls, and monitoring of blood hormone levels. As is standard, if there is any suspicion that the medications are being inappropriately administered, FSH assays will be checked.

Prescriptions, during this study, will be provided by the study nurse. Due to the nature of the protocol monitoring and the vigilance of these patients, prescriptions are not likely to be provided by a source outside of the institute and therefore contamination will be avoided. The physicians have been advised of the importance of protocol adherence and have agreed to abide by the randomized protocol. If contamination does occur, a per protocol analysis will be observed. The study and standard protocols differ significantly so that if a patient was to self medicate, this would be discovered on monitoring.

The nature of IVF treatment and monitoring requires strict observation of protocol (including timing of injection administration). Most of the patients to be included in this protocol will be patients who have previous experience with IVF and are familiar with its stringence. As is standard, patients will be advised to inquire about any concurrent

medications or procedures before proceeding. This includes a range from acupuncture to vitamins.

### **3.3.1.8 Sample Size and Recruitment**

Charts at the Cornell Institute for Reproductive Medicine will be reviewed to identify potential candidates. Only patients who have already planned to do an IVF cycle will be approached. Subjects will be approached by one of the principal investigators listed on this protocol after permission has been obtained from their primary physician.

Patients will be informed that participation is optional and in no way will their treatment be altered or compromised if they decline to participate. We expect to recruit approximately 30 patients per 3 month series and therefore the recruitment should be completed within a year. It is likely that recruitment will be completed in less than a year since we see approximately 600 IVF patients every 3 months and at least 30%(200) of these patients are poor responders by the above criteria.

An optimistic success rate for these patients is a 10% pregnancy rate. We feel a 30% (3-fold increase) success rate would represent an improvement. We arrived at this minimal clinically important difference by consultation with 2 of the pioneers and peer determined experts of the field. (Drs. Zev Rosenwaks and Owen Davis) Therefore with a beta of 0.20 and a 2-sided alpha of 0.05, our sample size requirement is 65 patients per group. (115)The drop out rate for these patients is expected to be extremely low (<1%).

However, there is a significant chance that they might have an elevated FSH level or E2 level that would prohibit them from starting medication. Therefore, the sample size was adjusted upward by 20% yielding a total sample size of 156 patients or 78 patients per group. An interim analysis will be done when approximately one third of the patients have been recruited to identify if there is a significant difference in treatment arms so as to prevent further enrollment of patients into a suboptimal treatment. Also, we will also assess recruitment on a monthly basis to ensure we are meeting a minimum of 10 patients per month. If recruitment is not adequate, we will reassess the methods of subject identification and study feasibility.

An interim analysis will be performed when approximately one third of the patients have been recruited. An alpha-spending sequential method (i.e. O'Brien –Flemming) is planned so that the amount of alpha used will depend on the fraction of total information available at the time of analysis. The analysis will be done at time = 1/3 because we expect to have 20% of the total sample size for whom final outcomes are not available concurrently enrolled. The final alpha for tests of significance for the primary outcome at the end of the trial will be adjusted accordingly.

The interim analysis is to serve the following functions:

- 5) to assess for any obvious superiority of one treatment versus the other;
- 6) to allow us to reassess the sample size based on the actual event rate from both groups to ensure the study has adequate power;
- 7) to allow us to check compliance with the protocol;
- 8) to allow us to reassess recruitment.

Indications for termination of the trial include inability to recruit the required sample size or obvious superiority of one treatment over another.

### **3.3.1.9 Analysis**

Data will be kept on all patients regarding age, parity, day 2 FSH and E2 levels. The primary outcome of pregnancy is a dichotomous variable and we will compare groups by the chi-squared test. Oocyte number, medication doses, days of stimulation, estradiol levels, fertilization rates, embryo number and quality, and number of embryos transferred will be compared by a student's t-test or Mann Whitney U test.

### **3.3.1.10 Data Collection, Quality control and Confidentiality.**

All patient data will be kept confidential as per strict policy at the Cornell Institute for Reproductive Medicine (CIRM). Stimulation, embryo and pregnancy results will be available to the patients primary physician and also to the principle investigators on this protocol. A copy of all study data will be kept in a confidential locked cabinet on the

CIRM premises and this data will be identifiable by code number. A logbook for codes will also be maintained under lock. However, patient's stimulation history will also be made a part of the patients medical record. Any computerized data will be available only to the principle investigators.

The Center has an excellent mechanism in place for follow-up of pregnancy data. The center is required by SART (Society for Assisted Reproductive Technology) legislation to report all outcomes of started IVF cycles. SART criteria for documentation of clinical pregnancy include ultrasound evidence of an intrauterine gestational sac. While most of the follow-up is done at our clinic until the patient is discharged (either with a negative day 28 bHCG (pregnancy test) or a positive fetal heart seen on ultrasound at 7 weeks gestational age or later), those patients who decide to follow-up with a physician closer to home will be required to provide us with documentation of outcome. Since 1985, only 10 patients have been lost to follow-up.

A training session will be held for all nurses, physicians, lab technicians, and administrative staff who might encounter these patients. Also, monthly luncheons will be held to update these same people about study progress and difficulties.

All assays will be run at the Center for Reproductive Medicine. If patients must have blood drawn at a different location, we will have that lab centrifuge and then freeze the specimen at -20 degrees Celsius and then FED -EX the specimen to us at our expense. This is routinely done for many of our out- of town IVF patients, so a procedure is already in place.

Since the study is not blinded with respect to the medications, error in prescribed versus taken drugs should not exist. The drugs will be prescribed by the study nurse or primary/co-investigators.

### **3.3.1.11 Economics**

An economic evaluation will be undertaken. Costs associated with the treatments (including any adverse event costs leading to the use of health care resources) and efficacy of the results of treatment will be compared in a cost-effectiveness analysis.

### **3.3.1.12 Avoidance of Bias**

The randomized nature of this trial will be the best tool against bias. Treatments will not be blinded but special attention will be given to concealed allocation.

### **3.3.1.13 Timeline**

With a recruitment rate of 10 patients per month, we expect this trial to be completed within one year.

## **3.3.2 Results of the Poor Responder Feasibility Study**

As noted in section 3.2.4, a feasibility study was proposed and conducted. The feasibility study followed the protocol for the poor responder study (section 3.5.2) with a reduced sample size. The sample size for the feasibility study was based on providing an overall estimate of the pregnancy rate. More specifically, if we estimate the pregnancy rate to be 10%, a sample size of 40 patients (2N) would give us precision of +/- 10%.

### **3.3.2.1 Study Outcomes**

The proposed study was approved by the Institutional review board in March, 2003. Recruitment began immediately thereafter and continued until April, 2004. During that time approximately 575 patients met the inclusion exclusion criteria. The principal investigator (SK) identified patients by review of more than 800 charts. The principal investigator then reviewed each of these charts with each subject's primary physician and approximately 114 subjects were identified as suitable candidates for the principal investigator to approach regarding study participation. Many subjects were deemed ineligible by their physicians for nonmedical reasons and at least 25% of potential subjects began ovarian stimulation for IVF before the primary physician could review for study suitability. Two patients were referred to the study by their primary physician and

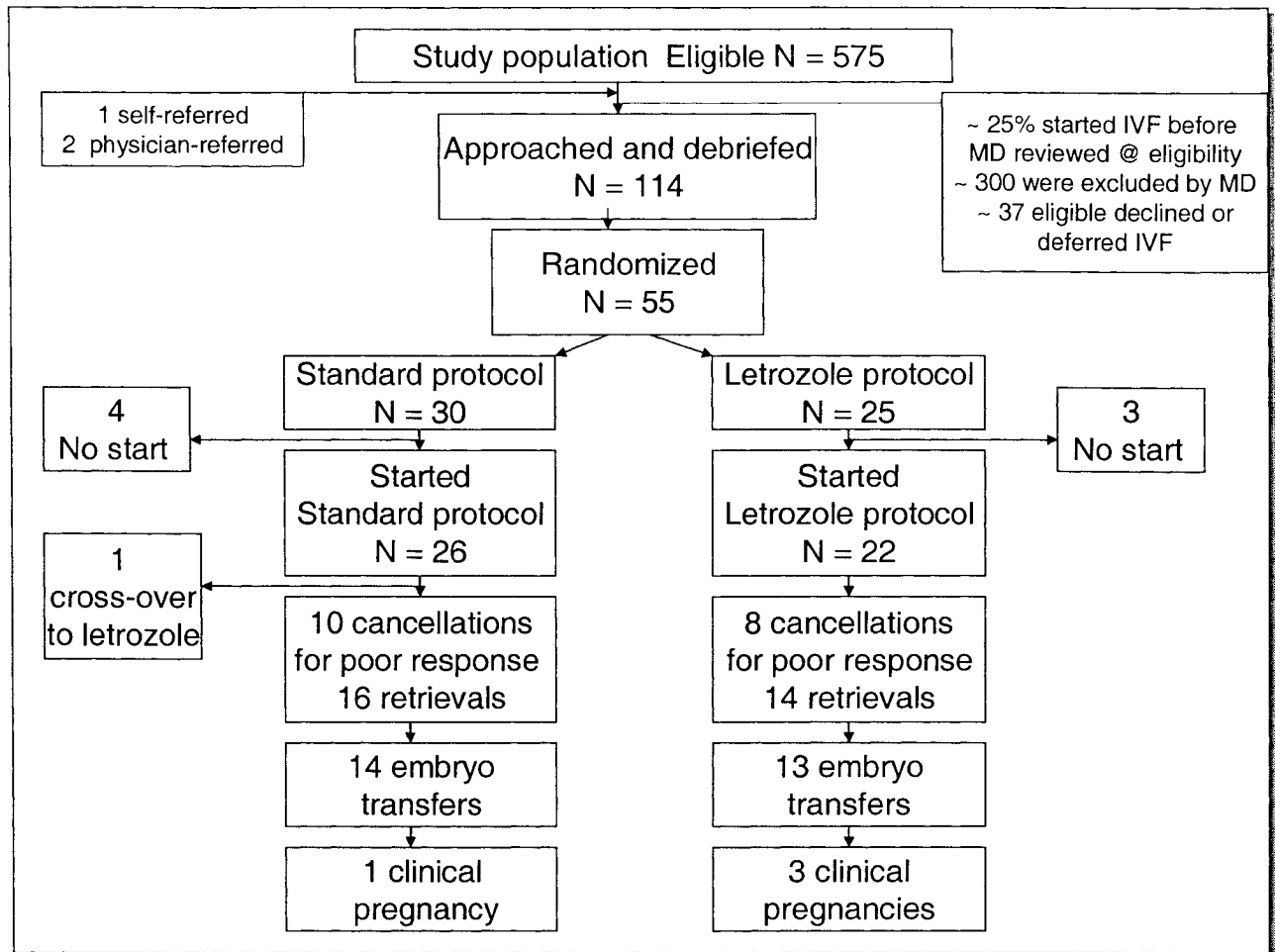
1 patient self-referred to the study after learning about it through a friend. Of 114 subjects, 59 agreed to participate but 2 declined participation on further review before randomization was initiated and 2 achieved a spontaneous pregnancy before randomization occurred. Six patients of 114 patients who underwent debriefing of the trial with the primary investigator declined participation but approached their primary physician to utilize the experimental protocol outside of the study and were prescribed the study protocol. While almost 50% of approached subjects agreed to participate, the majority of candidates indicated a preference for the study protocol but accepted participation because they were informed the experimental protocol was not available outside of the study. A database of patients from October, 2003 until April, 2004 identified 19 patients who were prescribed the experimental protocol outside of the study.

Of the 55 patients who were randomized, 6 patients were unable to begin ovarian stimulation for IVF because of elevated FSH or estradiol levels and one patient achieved a spontaneous pregnancy prior to ovarian stimulation and therefore did not start. This was anticipated and the recruitment adjusted. The percentage expected of such patients, who were randomized but not started on treatment, was 20%. The actual “no start” rate observed was 12.7% and these cases were equally distributed between the standard (12.9%) and letrozole (12.5%) protocol. All patients who began treatment were considered in the intention to treat analysis and followed until outcome. Cancelled cycles for poor response were included as outcomes.

One patient who was randomized to the standard protocol approached her physician to change to the experimental protocol. She crossed over to the experimental protocol but was analyzed according to the protocol to which she had been randomized.



**Figure 11: Flowchart of the randomized controlled trial.**



Patient demographics for each group are provided in table 7. Randomization was successful in equally allocating identified prognostic factors. There was some pressure from physicians to direct allocation but this was denied and as a result, some patients were prescribed the experimental protocol outside the study by their primary physician. There were no clinically or statistically significant differences between the groups. It is interesting that the mean number of inclusion criteria met in each group approached 3. The distribution of each inclusion criteria in both groups is found in table 8 and is also equally distributed in the two study groups. Most importantly, previously failed ovarian

stimulation, as indicated by previous E2 < 1000 pg/ml and previous oocyte yield less than 5, is considered the most important prognosticator of response to stimulation and is equally distributed between groups.

The Center for Reproductive Medicine and Infertility at Cornell University is an international referral center for patients suffering from infertility. Many subjects who have failed treatment elsewhere are referred to CRMI for CRMI's expertise and documented success in this clinically challenging area. As a result, there is some reluctance for both physicians and subjects to participate in RCTs. We, therefore, investigated the number of inclusion criteria met in those subjects who were permitted by their physicians to participate in this trial to determine if our study population was skewed. Table 9 indicates that over 63% of participants met 3 or more inclusion criteria. This predicts an even poorer prognosis than the original 10% pregnancy rate anticipated. The distribution of frequency of inclusion criteria met was equal in both groups.

Table 10 compares the characteristics of randomized subjects who did and did not start ovarian stimulation for IVF. There are no clinically significant differences. However, the standard deviation in the "no start" group for age and number of previous stimulations is large. We might expect a clinical difference in the characteristics of patients who started medication versus those who did not start. That is, those who did not start might have poorer prognostic factors. However, the number of "no starts" was equally distributed between the groups there were no striking difference between the randomized ineligible characteristics between the groups with the exception that one was a result of spontaneous pregnancy.

**Table 7: Poor responder patient demographics. Standard protocol versus letrozole protocol.**

Demographics	Standard protocol		Letrozole protocol	
	N = 30		N = 25	
	Mean	(SD)	Mean	(SD)
Age (years)	40.48	(3.25)	40.00	(3.52)
D2 <sup>a</sup> FSH <sup>b</sup> (mIU)	8.52	(5.84)	8.04	(2.97)
D2 E2 <sup>c</sup> (pg/ml)	40.32	(32.33)	38.07	(20.51)
Total number of previous cycles	2.37	(2.18)	2.71	(2.12)
Number of CRMI <sup>d</sup> previous cycles	1.35	(1.23)	1.21	(1.41)
Gravida	1.23	(1.20)	1.54	(1.40)
Number of inclusion criteria met	2.81	(1.30)	2.92	(1.25)

<sup>a</sup>D2 – day 2 of menses

<sup>b</sup>FSH – follicular stimulating hormone

<sup>c</sup>E2 - estradiol

<sup>d</sup>CRMI – Center for Reproductive Medicine and Infertility at Cornell University

**Table 8: Frequency of each inclusion criteria in overall, standard and letrozole groups.**

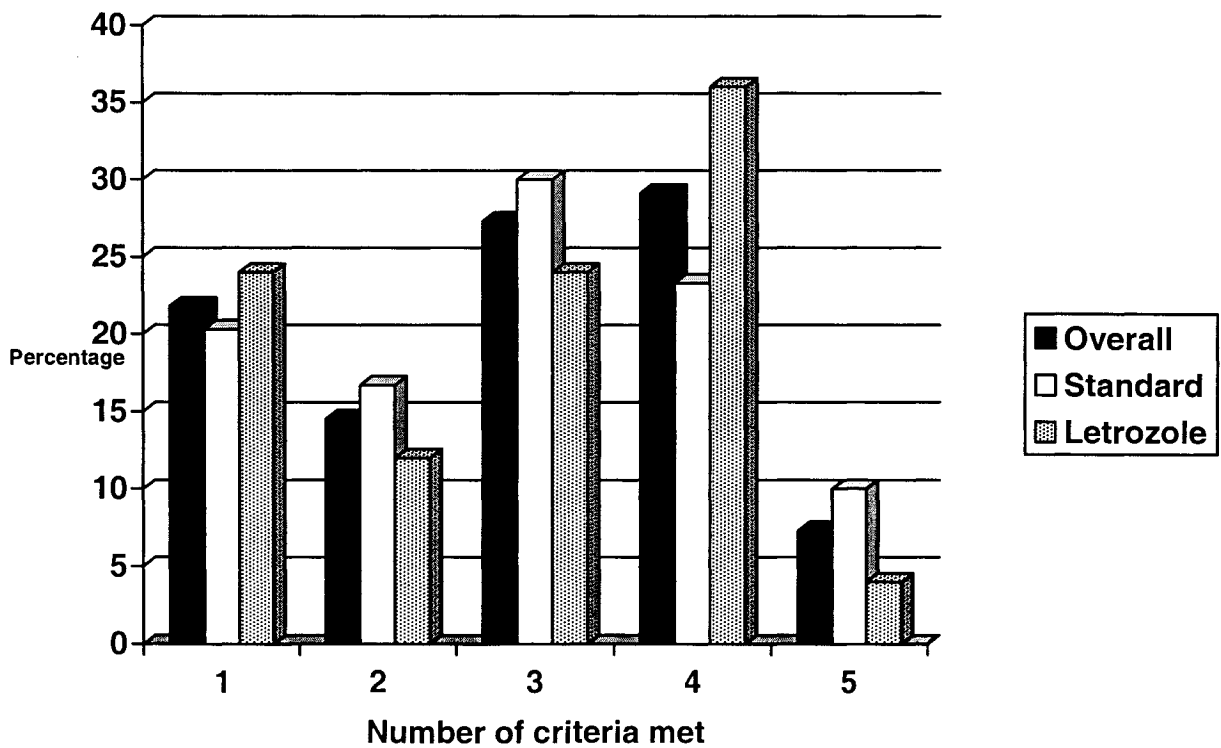
<b>Inclusion criteria</b>	<b>Overall (N = 55) %</b>	<b>Standard protocol N = 30 %</b>	<b>Letrozole protocol N = 25 %</b>
<b>Age (&gt;40 years)</b>	65.5	70.0	60.5
<b>Previous cancelled stimulation</b>	55.6	53.3	56.0
<b>Previous stimulated peak E2<sup>a</sup> &lt; 1000 pg/ml</b>	63.6	63.3	64.0
<b>Previous IVF oocyte yield &lt; 5</b>	50.9	50.0	52.0
<b>Previous high FSH<sup>b</sup> / E2</b>	47.3	46.7	48.0

<sup>a</sup>E2 –estradiol; <sup>b</sup>FSH – follicular stimulating hormone

**Table 9: Frequency of participants who met greater than 1 inclusion criteria.**

	<b>Overall</b>	<b>Standard Protocol</b>	<b>Letrozole Protocol</b>
	<b>N = 55</b>	<b>N = 30</b>	<b>N = 25</b>
<b>Number of inclusion criteria met</b>	<b>%</b>	<b>%</b>	<b>%</b>
1	21.8	20.3	24.0
2	14.5	16.7	12.0
3	27.3	30.0	24.0
4	29.1	23.3	36.0
5	7.3	10.0	4.0

**Figure 12: Frequency of number of inclusion criteria met**



**Table 10 : Comparison of patients who were randomized but did not start stimulation versus patients who did start stimulation. (overall- standard and letrozole groups together)**

	<b>Stimulations started</b>		<b>Stimulations not started</b>	
	<b>N = 48</b>		<b>N= 7</b>	
	<b>Mean</b>	<b>(sd)</b>	<b>Mean</b>	<b>(sd)</b>
<b>Age (years)</b>	40.54	(2.91)	39.17	(5.64)
<b>Gravidity</b>	1.38	(1.32)	1.50	(1.38)
<b>Number of inclusion criteria met</b>	2.85	(1.29)	3.00	(1.26)
<b>Total previous number of IVF</b>	2.60	(1.96)	1.83	(3.55)

Note: Standard protocol (N = 30) had 4 no starts and letrozole protocol (N = 25) had 3 no starts.

**Table 11 : Clinical pregnancy rates**

	Standard protocol		Letrozole protocol		p-value	RR	95% CI
	N = 26		N = 22				
	Frequency	%	Frequency	%			
<b>Clinical pregnancy rate per cycle started***</b>	1 / 26	3.8	3 / 22	13.6	0.3055	3.86	0.43, 34.49
<b>Clinical pregnancy rate per retrieval</b>	1 / 16	5.9	3 / 14	23.0	0.1912	3.69	0.43, 31.43
<b>Clinical pregnancy rate per transfer</b>	1 / 14	6.7	3 / 13	25.0	0.2442	3.23	0.35, 27.38
<b>Multiple pregnancy rate/observed pregnancy</b>	1 / 1	100.0	1 / 3	33.3	0.8194	0.25	0.05, 1.36
<b>Implantation rate (# gestational sacs/ total number of embryos transferred)</b>	2 / 33	6.1	4 / 42	9.5	0.5831	1.57	0.31, 8.06

RR = relative risk; CI = confidence interval

Clinical pregnancy as per SART = presence of 1 or more gestational sacs

Note: 1 twin pregnancy observed in standard and letrozole group each.

\*\*\*primary outcome on which sample size is based

The primary outcome assessed is clinical pregnancy rate per cycle started. (Table 11)

This is the measure that is also used by the Society for Assisted Reproductive

Technology (SART) to report clinic success rates. SART defines clinical pregnancy as

the presence of one or more gestational sacs on ultrasound. A chi-square test with a 2-

tailed p value for equality of proportions was employed to compare groups. The observed pregnancy rate in the standard group is 3.8%, less than half the 10% pregnancy rate expected. The lower observed pregnancy rate reiterates the skewed nature of the subset of the study population we obtained by including women who met more than 1 inclusion criteria. The pregnancy rate in the letrozole group was 13.6% (p=0.3066). This represents a more than 3-fold difference between pregnancy rates. The numbers for the feasibility study are small and the results may be reversed in a larger study but a clear trend is observed here.

The trend continues for the other clinical pregnancy rates observed. Most importantly, clinical pregnancy rate per transfer is often considered the mostly clinically sound measure of a given institution's success rates. The observed clinical pregnancy rate per transfer in the study group was 6.7% versus 25% in the letrozole group (95% CI 0.35, 27.38). The implantation rate (IR) does not show the same strength of association between the groups (6.1 versus 9.5%).

Secondary outcomes were analyzed and p-values documented in Table 12 illustrate a trend towards a statistically significant difference for those outcomes which are clinically significantly different.

There were no appreciable differences in cancellations rate, numbers of retrievals or transfers per started cycle, or number of transfers per retrieval between the standard and experimental protocol. There was no difference between doses of medications (FSH and hMG) used between the groups. All embryos were transferred on day 3 and no excess embryos were available for cryopreservation. There was no difference in embryo grade or endometrial thickness at the time of transfer between the groups despite lower estradiol levels in the letrozole group. Patients in the letrozole group experience a lower E2 on day 6 of stimulation, as expected. E2 levels did not start to rise until 2 days after discontinuation of letrozole (day 8) but this did not appear to affect endometrial thickness.



**Table 12: Secondary outcomes in subjects who began treatment.**

Outcome variable	Standard protocol		Letrozole protocol		p-value
	N = 26		N = 22		
Cancellation for poor response before hCG <sup>1</sup>	10	(38.4 %)	8	(36.4%)	NS
Retrievals/started cycle	16/26	(61.5%)	14/22	(63.6%)	NS
Transfers/started cycle	14/26	(53.8%)	13/22	(59.1%)	NS
Transfers/retrieval	14/16	(87.5%)	13/14	(92.8%)	NS
E2 <sup>2</sup> day of hCG (pg/ml)	833.70	(369.40)	666.52	(310.19)	NS*
E2 day post hCG (pg/ml)	1058.06	(438.39)	795.29	(473.76)	NS*
Total FSH <sup>3</sup> (IU)	2174.35	(835.80)	2198.68	(984.94)	NS
Total hMG <sup>4</sup> (IU)	1575.00	(835.17)	1366.68	(482.98)	NS
Day of hCG	11.00	(1.76)	11.63	(3.06)	NS
Oocyte number	4.31	(2.21)	6.46	(3.44)	0.052*
Immature oocytes	1.25	(1.69)	1.71	(2.64)	NS
ICSI <sup>5</sup>	14/16	(87.5%)	11/14	(78.6%)	NS
Fertilization rates	2.33	(1.23)	3.50	(1.45)	0.027*
2PN <sup>6</sup>	2.25	(1.23)	3.43	(1.40)	0.021*
Other PN <sup>^</sup>					
Number embryos transferred	2.36	(1.08)	3.23	(1.30)	0.069*
Grade	2.27	(0.70)	1.96	(0.92)	NS
Endometrial thickness (mm)	9.82	(2.16)	9.64	(2.21)	NS
Number embryos cryopreserved		0		0	NS

<sup>1</sup>hCG - human chorionic gonadotropin; <sup>2</sup>E2—estradiol; <sup>3</sup>FSH- follicular stimulating hormone; <sup>4</sup>hMG- human menopausal gonadotropin; <sup>5</sup>ICSI -intracytoplasmic sperm injection; <sup>6</sup>PN -pronuclei

\* clinically important differences; ^ not estimable

p based on 2-tailed t-test or chi-square as appropriate

There was a trend towards a lower E2 level in the letrozole versus the standard protocol groups for both the day of and after hCG. In an unstimulated cycle, a mature follicle is usually associated with an estradiol level of 200-250 pg/ml. The E2 level in the letrozole group is approximately 200-250 pg/ml less than in the standard group, however, the standard deviation is wide. The physiology of letrozole, described earlier, predicted this outcome. This lower E2 level, however, did not appear to affect endometrial thickness. Interestingly, despite a lower E2 level in the letrozole group, there was a clear trend towards a higher oocyte number, fertilization number, 2PN number and number of embryos transferred. The fertilization number does not represent the percentage of oocytes successfully fertilized in each group but rather the average, absolute number of fertilized pre-embryos available. The egg (oocyte) yield in the standard group was 4.31 and in the letrozole group was 6.46. The egg yield in the letrozole group was on average 2 eggs more than the standard group and was, on average, greater than 5. Less than 5 oocytes retrieved at the time of ovarian stimulation classified a subject as a poor responder and the letrozole protocol appears to have made a significant impact on this prognostic factor. The extra oocyte yield in the letrozole group resulted in an average of 1 more oocyte fertilizing and reaching the 2PN stage in the letrozole group versus the standard group. Therefore, an average of 1 more embryo was transferred into the uterus in the letrozole group versus the standard group. (3.23 versus 2.36 ET) An extra embryo available for transfer in any IVF patient and particularly a poor responder is a very important difference clinically.

Patient demographics at the start of this study were equally distributed between groups. Nevertheless, response to stimulation appeared clinically better in the letrozole group than in the standard group. Again, the numbers of this feasibility study are small.

Characteristics of the pregnancies achieved are represented in Table 13. Three of four pregnancies were achieved in the letrozole group. All subjects met 2 or more inclusion criteria and all subjects had undergone previous failed ovarian stimulation for IVF elsewhere. Only one subject had had a previous IVF cycle at CRMI. The three subjects who conceived with the letrozole protocol were all under the age of 40 years (average age

of both groups) despite the fact they met at least 2 of the inclusion criteria. Therefore age still appears to be the one of, if not the most important prognostic factors for achievement of pregnancy.

### **3.3.2.2 Feasibility Outcomes**

The objectives of the feasibility study were 4-fold: to determine a precise estimate of the actual pregnancy rate in the recruited control group, to determine the ability to randomize subject and conceal allocation and to recruit the number and type of patients required for the trial.

The observed pregnancy rate was less than half of the expected pregnancy rate (3.8 vs 10%).

The methods used to randomize were effective as evidenced by the comparability of groups in table 1. These included an independently computer-generated sequence of random numbers in blocks of 4 to 6. Sealed opaque, sequentially numbered randomization envelopes were then created by a third party and allocation was performed by the research nurse. Pressure to subvert allocation was significant but prevented by concealment of allocation. Absence of concealed allocation in this situation almost certainly would have led to direction of allocation. Nevertheless, a significant number of patients began the experimental therapy outside of the study because they refused to be allocated in a concealed fashion. We were not successful in avoiding use of the study protocol outside the study.

We were also not successful in maintaining physician enthusiasm for the study over time. Recruitment peaked during the first 3 months. Within a year, we were able to recruit the sample size for the pilot study. However, the type of patients we recruited were skewed towards the worst prognosis group, as evidenced by the fact that more than 64% of subjects met three or more of the inclusion criteria rather than just one. Also, patients who conceived with a previous protocol were excluded. Only 1 patient was directly referred to the study by their physician and 2 patients self-referred. Our subject

identification process was not very efficient; more than 800 charts were screened and of 575 eligible patients, 114 were approached and 55 consented. Uptake for the study was high, however, and approximately 50% of debriefed subjects consented to participate.

An assessment of randomized ineligibles suggested the incidence was lower than expected (13% vs 20%). Patients did not appear to be cancelled from during ovarian stimulation more often or earlier in one group than another which also suggests clinical equipoise.

Protocol	Age	Number of Inclusion criteria met	Total # previous cycles	Total # previous CRMI <sup>a</sup> cycles	Day 2 FSH <sup>b</sup> (pg/ml)	Day 2 E2 <sup>c</sup> (pg/ml)	Oocyte number	Number of embryos transferred	Grade	Outcome	Day of hCG <sup>d</sup>	Peak E2 (pg/ml)
Standard	44	4	1	1	4	43	4	3	2	2 sacs	12	523
Letrozole	38	3	2	0	1	<20	17	3	2	2 sacs	10	1161
Letrozole	36	2	3	0	3	<20	5	3	4	1 sac	13	848
Letrozole	37	3	2	0	9	21	6	4	2	1 sac	12	294

<sup>a</sup>CRMI – Center for Reproductive Medicine and Infertility at Cornell University

<sup>b</sup>FSH – follicular stimulating hormone

<sup>c</sup>E2 – estradiol

<sup>d</sup>hCG – human chorionic gonadotropin

**Table 13: Characteristics of subjects who achieved pregnancy**

### **3.3.3 Discussion**

#### **3.3.3.1 Discussion of Results of the Pilot Study**

The results of the pilot study are extremely interesting and very encouraging. While the numbers for this study are small and there are observed trends towards a clinically important difference in pregnancy rates, caution must be stressed since these results may reverse in a larger trial. Nevertheless, we feel a larger trial is warranted.

The observed clinical pregnancy rate per cycle started was considerably less than expected (3.8% versus 10%). One reason for this may be that patients who were poor responders but who had achieved pregnancy with a previous stimulation protocol were excluded from the study. We therefore excluded perhaps the best prognosis subset of this group. Nevertheless, at the CRMI, a patient who experiences a successful outcome with a particular protocol automatically repeats that protocol if a subsequent treatment cycle is undertaken, in order to maximize subsequent success potential. It would be unacceptable to patients and physicians to alter this policy. Secondly, our data demonstrated that the subset of patients who were permitted to participate in this trial were skewed towards the worst prognosis group as over 63% met 3 or more inclusion criteria. These patients would be denied treatment in most other IVF institutions. We collected information as to the next treatment step for each of our subjects after they completed our study. The majority of patients discontinued therapy, had one more unsuccessful attempt, or subsequently pursued donor egg options. One patient conceived on a subsequent cycle. These circumstances make our findings even more interesting.

Interestingly, the percentage of cancellations, the percentage of cycles that proceeded to retrieval and the percentage that proceeded to transfer were not different between the groups. Nevertheless, the clinical pregnancy rate per transfer was even more striking in the letrozole group versus the standard group (6.7% versus 25%) than the clinical pregnancy rate per cycle started. However, the implantation rate (number of gestational sacs observed/ total number of embryos transferred) did not appear to be different. (letrozole group 9% and standard group 6.7%). The secondary outcomes serve as useful

surrogate markers to explain this phenomenon. Despite the fact that the estradiol levels were lower in the letrozole group than the standard group, the number of oocytes retrieved was, on average, 2 more per case than the standard group (4.31 versus 6.46). This resulted in more fertilized preembryos in the letrozole group (3.50 versus 2.33) and more embryos available for transfer (3.23 versus 2.36). One of the inclusion criteria for poor response was a previous history of less than 5 oocytes retrieved. Therefore, a mean of 6.46 oocytes in the experimental group versus the standard group is a very important clinical difference.

Also, an additional embryo available for transfer is an extremely important clinical difference in any IVF patient group and even more so in the poor responder group. Given that the implantation rate was not significantly different, the observed difference in pregnancy outcomes may be related wholly to the number of embryos available for transfer and the fact that this number was higher in the letrozole group. Again, however, the differences in implantation rate may not be adequately reflected due to the numbers in this pilot study. The implantation rates may change significantly in a larger study. Lower estradiol levels in the letrozole group did not appear to compromise endometrial thickness, oocyte yields, or pregnancy rates.

The characteristics of the patients who achieved pregnancy reveal that age is still the most important overall prognostic factor for achievement of pregnancy, even in younger patients who have a document history of low response to stimulation. This is expected as aging oocytes are known to have meiotic spindle dysfunction and therefore result in genetically abnormal embryos more frequently which usually do not implant or end in biochemical pregnancy or miscarriage. Increasing the number of oocytes and therefore embryos available for transfer increases the probability of having a normal embryo which may attain viability.

The results of the pilot study should be pursued in a larger trial. However, if these trends prove true, letrozole may have some benefit in influencing the quality of stimulation in these women who have diminished ovarian reserve. These benefits may relate to the the

earlier stated theories and speculation regarding the role of androgens in follicular development.

### **3.3.3.2 Assessment of Feasibility Study Objectives**

The objectives of the feasibility study were 4 fold. Firstly, identify issues in types and numbers of patient recruitment. Secondly, to ensure that randomization can be accomplished. Since, treatments are not blinded, it is essential that treatments be allocated randomly. Next, we wanted to gain accuracy and precision around the expected pregnancy rate in the group of patients being recruited for the study. We anticipated a skew in the subset of the population that would be accessible for the RCT and therefore predicted a lower than expected pregnancy rate of 10%. Finally, we endeavored to run the pilot study in order to identify logistic problems before embarking upon a full protocol.

#### **3.3.3.2.1 Recruitment**

The first objective was to establish whether recruitment was feasible. There are two issues regarding recruitment. We estimated a lower standard pregnancy rate for the clinical trial than the expected pregnancy rate for the general group of poor responders. We lowered the estimate because we expected there would be some skewness towards an even poorer prognosis in the subset of the population that we would be permitted to randomize. Therefore, one of the issues of recruitment is whether we are able to recruit the type of patient (inclusion/exclusion criteria) we intend. The second issue is whether we can recruit the numbers of patients required.

Clearly, as the distribution of inclusion criteria demonstrate, there is a significant skew to a subset of worst prognosis patients. This is likely due to the fact that there is some desperation for both the physician and patient in these cases and this desperation factor encourages them to participate in an experimental protocol and trial. This is further reiterated by the lower pregnancy rates and the fact that most patients proceeded to termination of all therapy or donor egg after their study cycle. The skewness of the population is a problem for two reasons. First, many or all of these patients had been or would be denied treatment at other IVF centers. Indeed, most centers would not accept



patients who meet 2 or more of our inclusion criteria. This fact influences the generalizability of our results. Even if a larger trial proves the experimental therapy to be beneficial, it is unlikely to be applied to the same subset of patients at different institutions. We have two options; we may either accept the subset presented as the group which we can access for such a trial or else we must develop a strategy to target a wider section of this population. One of the strategies for the latter may be to include women who had previously achieved pregnancy with IVF, despite their history of poor response. Another strategy may be to score the inclusion criteria according to prognostic value and only accept patients who meet less than a predetermined maximum score or number of inclusion criteria. Another alternative may be to include patients who are referred for their first cycle at CRMI after having failed cycles elsewhere as opposed to including patients who have already attempted cycles at CRMI. There are several problems in trying to increase the breadth of our sample population. The main problem is that reluctance to participate in RCTs even in this group. Such reluctance and pressure to direct allocation of study treatments resulted in at least 19 patients being prescribed the letrozole protocol outside of the study. The best way to increase the breadth of the sample population may be to open the study to other centers where the policies are not so strict and the patient population not so difficult. However, the heterogeneity introduced by different ovarian stimulation management styles and expertise of the IVF lab would be a significant consideration. At the CMRI, the final decisions are made by the director and chairman of the department. Each of the attending physicians are trained by the CRMI. Therefore the decision making process, although tailored to each individual patient case, is consistent. By introducing a multi-center trial, heterogeneity by intersite variability (different centers) as well as by individual physician management styles (intrasite variability) will become an issue. Our second option is to accept the population to which we have access then extrapolate, on physiologic principles, that the benefit may extend to subjects with a milder form of poor response. If we are able to demonstrate a clear benefit of such an RCT in the accessible population, we may be in a better position to mount similar studies in a more generalizable group at a later date.

The second issue in recruitment is whether we can obtain the required number for a larger trial. There was a significant “funnel effect” from the eligible population to the randomized population. We observed that enthusiasm for the study was highest in the first 3 months of the trial. As time progressed, other novel therapies were proposed, and physicians may have become frustrated with not being able to obtain preliminary study results. During the year our study was conducted, another novel therapy was introduced for the same group of patients. That therapy was not subject to randomization and physicians began prescribing it to patients who were eligible for our study. To complicate matters, an uncontrolled case-series analysis suggested that the pregnancy rate with the newer therapy was almost 30%. Although this evidence was essentially anecdotal and potentially highly biased, it was strong enough to convince physicians to direct treatment.

Another obstacle in recruitment was that at least 25% of potential candidates were lost because they started treatment after the principal investigator (PI) had established their eligibility but before the primary physician had reviewed the chart and given permission for them to be approached. Only 2 patients were referred to the study by the primary physician themselves. One patient referred herself after hearing about the trial from a co-patient. Therefore, recruitment was almost entirely a result of the principal investigator’s efforts to actively search for patients. Paper charts were required for the principal investigator to establish subject eligibility. Charts were often not centralized and difficult to locate. Attainment of charts in a timely fashion was a significant obstacle. The rate-limiting step, however, was the inability to have eligible charts reviewed by the primary physicians in a timely manner.

No patient was permitted to be approached without the consent of the primary physician. It would take the physician, on average, 30 – 60 seconds to review a chart for suitability. However, the increased work load was still onerous for physicians. This resulted in only 20% of eligible subjects being debriefed and therefore only 10% of eligible subjects recruited despite a 50% patient acceptance rate. Unfortunately, much effort was wasted in this way, with multiple cancelled appointments, pulling and repulling of charts and many potential candidates were lost because they started stimulation before

permission was obtained to approach them. A more efficient strategy and stronger physician commitment would be required in order to proceed with a larger trial.

The Institutional Review Board at New York Presbyterian Hospital requires that the primary or co-investigators recruit and consent study participants. This is contradictory to many stated practices which try to limit coercion of subjects. However, a valuable recruitment lesson was learned from this obligated practice. Prior to this study, the CRMI experienced difficulty to actively recruit more than 3- 5 patients for any prospective study (randomized or not randomized) requiring informed consent. Fifty percent of approached patients agreed to participate in this study. The uptake could well have been higher, however, the principal investigator was very wary of issues of coercion. Therefore, the recruitment process involved a 45 minutes debriefing appointment between the subject and PI. Even if patients were initially enthusiastic to participate, the PI declined enrollment until another appointment was made after the subject and/or her partner had reviewed the consent form in order to answer any arising questions. Potential subjects were also given a direct line to the PI which they could use at any time to answer questions that arose. Because of concerns about coercion, any subject that was less than completely comfortable with the study process was advised not to enroll. The entire recruitment process required on average 3 hours of discussion with each potential candidate before a decision was made.

Some of the positive influences in patient uptake included the fact that prognosis was poor, the protocol was experimental and not supposed to be available outside the study. All patients were assured that their primary physician had reviewed their chart and considered them eligible for the study because there was equipoise regarding treatment alternatives, and patients were reassured that study participation involved randomization of the starting protocol but daily monitoring and medical management decisions would proceed unaltered and the primary physician, in combination with the director, would still be responsible for their care. The research nurse involved with this study had been involved with previous studies and stated that she felt the difference in uptake was largely due to the fact that the PI (person obtaining consent) was an infertility trained medical

doctor who could answer patients other questions immediately rather than a non-MD who would have to tell these patients she would request a physician to get back to them, which could then take a long time and frustrate patients. Interestingly, a major concern for patients who are considering participating in a trial is that they will receive substandard or impersonal care. An official survey was not conducted on these patients, but most spontaneously commented to their primary physician and/or the study nurse that they felt more attended to in the study and that was a large reason they agreed to participate.

Therefore strategies to enhance recruitment may include the following: to have an MD/ highly educated/informed individual responsible for debriefing the patients; to hire at least one fulltime person responsible for patient identification and recruitment (the PI in this case was a fulltime clinical fellow); to encourage or offer incentive to the file room staff to provide the recruitment coordinator with charts in a timely fashion; to offer an incentive to physicians not to recruit patients but merely to honor timely appointments with the PI to review patient charts; to develop a strategy to more efficiently identify eligible candidates (ie enlist nursing, resident and fellow medical staff to remind physicians of the study and identify patients to the PI), and perhaps to involve patients in recruitment by posting advertisements that encourage them to ask and therefore remind their physician about the study. It was brought to our attention that this study was being discussed in patient web chat rooms and we had several patients from other centres call the PI to inquire about the protocol. Finally, addition of multiple centres may need to be added if recruitment suffers but the analysis must account for possible sources of heterogeneity.

#### **3.3.3.2.2 Randomization**

The second objective of the feasibility study was to determine if randomization was possible and concealment of allocation achievable in this group of patients. Our demographics table in the results section illustrates that randomization was successful and known prognostic factors and inclusion criteria were equally distributed between study groups. However, there were several potential obstacles to randomization. As expected, there was substantial pressure to subvert the randomization process and direct allocation. On several occasions we received pressure to “open the next envelope to see if it

contained the desired treatment” before enrolling subjects. By declining, however, we did lose several potential subjects who were then assigned the experimental protocol outside the study. Such “deciphering” of allocation could have clearly impacted our results and concealment of randomization allocation was especially important since treatments were not blinded. In order to prevent “deciphering”, requests to direct allocation were denied, computer generated randomization lists incorporated blocks that varied from 6 to 4 which, in addition to stratification of standard protocols made deciphering extremely difficult, and allocation codes were kept in sequentially numbered opaque sealed envelopes to which only the study nurse and PI had access. The randomization list was computer generated by an independent agent based on sample size and stratification information. Only the research nurse, not the PI or the physicians, had access to the randomization list and she was responsible for completing the allocation envelopes. The importance of randomization is demonstrated by equal distribution of prognostic factors between treatment groups. While some authors and clinicians suggest that equality of groups may be attained by a “matching process”, such a process may be accompanied by considerable bias and does not ensure equal distribution of unknown prognostic factors. While the randomized controlled trial is counterintuitive to many physicians and “appears to annoy human nature – if properly conducted, indeed they should.” Indeed, proper, concealed randomization is the best way to avoid bias and is especially important in avoiding confounding and achieving equal distribution of potentially unrecognized prognostic factors. Issues associated with randomization are discussed in referenced papers.(116, 117) The methods we employed in this feasibility study were successful and we would advocate repeating them. Avoidance of temptation by not allowing access to the randomization list or envelopes was, perhaps, one of the strongest mechanisms to prevent deciphering. Further, a central randomization process would also help ensure concealment of allocation, especially in a multicentre study.

#### **3.3.3.2.3 Determination of the primary outcome variable and sample size for the definitive trial**

The primary outcome variable was clinical pregnancy per cycle (patient) started. Clinical pregnancy was defined according to SART guidelines as the presence of gestational sac

observed on ultrasound. Clinical pregnancy is rarely, if ever, used as the primary outcome variable in studies of reproductive medicine. More common are assessments of surrogate outcomes such as ovarian response to stimulation as measured by estradiol levels, oocyte numbers and sometimes embryo numbers. Where pregnancy rates are considered as outcomes, day 28 bhCG levels are usually reported since the attrition rate by the time of ultrasound is significant. Day 28 bhCG levels are not considered evidence of pregnancy by SART because a significant proportion may result in biochemical pregnancies which resolve before a clinical pregnancy is established. Clinical pregnancy, as defined here, is clearly a hard outcome and not subject to biased interpretation. Biochemical pregnancies are softer outcomes. We recommend maintaining “clinical pregnancy” as the primary outcome of the full protocol.

The observed pregnancy rate for the group of patients defined by the preset inclusion-exclusion criteria at CRMI may approach 30%. This pregnancy rate is much higher than the 10% or less achieved at other centres. However, we estimated a pregnancy rate of 10% based on the fact we expected to enroll subjects who were at the lower prognostic end of the group. The observed pregnancy rate per cycle started was 3.8% - less than half that expected. The distribution of inclusion criteria, as observed in the results section, predicts this outcome. The observed pregnancy rate in the experimental group was 13.6% per cycle started, almost 4 fold the standard pregnancy rate. However, if we use clinical pregnancy per transfer, which is thought by many institutions to be a better marker of success, the difference is larger (6.7 versus 25%) and the final sample size for the definitive trial would be smaller. Nevertheless, we believe that the important clinical question that most physicians and patients want to know is: What is the success rate ie pregnancy rate going to be if a patient begins a particular protocol, taking into account the potential failures at each step? Therefore, we resolve to maintain clinical pregnancy rate per cycle started as the primary outcome. Based on observed incidences, a sample size calculation incorporating a power of 80% and alpha of 0.05 would require 131 patients per group to definitively conclude a statistically important difference. On a positive note, the “no start” rate was less than expected at 12.7 instead of 20%. The sample size may be adjusted accordingly.

In the pilot study, we did not assign alpha values to secondary outcomes. Nonetheless, we did observe clinically impressive differences in secondary outcomes which supported the direction of the trend of the primary outcome. For the definitive trial, we have decided not to spend alpha on the secondary outcomes but simply to qualitatively evaluate these outcomes as part of the interpretation of the primary findings.

#### **3.3.3.2.4 Logistic issues in trial implementation**

Issues of recruitment and randomization have been addressed earlier. A major problem was timely and efficient attainment of permission from primary physicians to approach eligible subjects. An incentive mechanism to encourage timely review of such subjects, not to actually recruit patients, may help. A second issue was the frustration and eventual ambivalence of the physician group toward the experimental protocol because results were not available to them. The best technique against this may simply be education and evidence from this trial that completing the pilot was valuable. If results had been opened earlier on, the clear direction of trend would not have been observed and the study may have been terminated for recruitment issues. A third issue was the negative attitude of the physicians towards RCTs as evidenced by the use of the experimental protocol outside the study, cross-over of one subject, and pressure to direct allocation. Unfortunately, this obstacle is much harder to overcome. Only through demonstration of value and education of the physician group by example will this be overcome. In this particular circumstance, leadership and regulation by the director in terms of enforcing adherence to trial protocols is necessary to avoid such behavior. While cross-over was an issue in one case, uninformed contamination was not and this was affirmed by the fact that estradiol levels were monitored frequently in both groups and as such would have demasked contamination and knowledge of letrozole was necessary for clinical management. Also, while this group of patients tend to be very aggressive regarding pursuing therapy, they are also extremely compliant with prescribed therapy and adhere to instructions strictly for fear of ruining a stimulation cycle. Nevertheless, we may not be able to significantly influence these latter issues in the short term.

The pilot study was a “pragmatic” trial. The physicians prescribed an initial protocol of their choice (between 2 standard poor responder protocols) and permitted certain patients to participate in the trial. Randomization of the prescribed versus study protocols was the only way in which management differed from study and nonstudy patients. Allocation was concealed but treatment was not blinded. The clinical management was decided in the usual fashion on a day to day basis. This was the only way that patients and physicians would accept this trial and practically makes sense since this is how patients at this centre would be managed outside the trial. At this center, the approach is justified since the medical management is consistent and the final decision is made by one person. The highest standard of care is applied to each individual. In a different situation, such as a multicentre trial, a more explicit management protocol will be necessary to avoid inter and intra centre variability. If a multicentre trial is initiated, we will need consistent protocols or central clinical management in order to mimic the one centre/one person rule and maximize outcomes. However, the clinical management flexibility issues detracted from our PCOS RCT and the flexibility here for physicians to make clinically indicated decisions facilitated the study process regarding recruitment and adherence.

Sources of bias were limited in this study by the efficient randomization process and the fact that the results have been contained from the physician group. However, one potential source of bias was identified. Because treatments were not blinded, physicians may have had more or less of an inclination to cancel patients depending on their personal bias about the utility of a given treatment. That is, if some physicians really did not believe in the letrozole protocol, for example, they may have cancelled patients earlier than had patients been on the standard protocol. While this could be a potential problem, cancellation rates did not vary between groups in the pilot study. Since most of the subset of patients we recruited were on their last attempt, the bias was generally to push them as far as possible, whichever the allocated group. It appears to us that patients were treated equally in this way.

Issues of patient selection are major. We must decide if we proceed with the subset we have isolated or if there is a way to widen the study population. The inclusion –exclusion



criteria are sufficiently wide to include less severe poor responders. One way to achieve this may be to limit the number of inclusion criteria a subject can meet or else to score the inclusion criteria according to importance and have a maximum attainable score for patients allowed to participate. This may however, simply reduce our accessible population. Another option is to open the study to a multicenter trial. The severity of poor responders at other centers is likely considerably less and we could still employ a scoring system to characterize candidates. However, this trial is very similar in principle to a surgical RCT. The techniques, skill, and level of attention to detail vary widely from center to center. There is a substantial body of evidence that surgical outcomes improve with the volume of procedures done. The Society for Assisted Reproductive Technology publishes outcomes data on each US IVF center. It has been documented that larger volume centers (generally > 1000 cycles per year) have better outcomes.(118) Most clinics average 200-300 cycles per year. A larger center is considered 500-600 cycles. CRMI starts approximately 2500 IVF cycles per year. The advantage we have at the CRMI is that it is recognized as the gold-standard IVF institute because of attention to detail, highly sophisticated lab expertise, unsurpassable success rates in the most difficult cases and the fact that it essentially functions as an autocracy with the director/ chairman (same person) running an extremely tight organization and being involved in the micromanagement of each step and decision. Therefore, it is the one place where we can be assured that control of every other potential factor which may come into play from the medical management and institution side is controlled equally in both groups. The alternative to a multicenter trial is to continue the study in this institution with the given population where the results will be valid and later extrapolate to a more generalizable population.

During the course of this trial, a more novel therapy was proposed. Physicians began applying it to patients who were eligible for our study in a case series fashion. Just prior to completion of the pilot study (March 2004), an early analysis of the case series data suggested that the protocol was of benefit compared to previous protocols in the same patients as their own historical control. The pregnancy rates approached 30% (the same as the pregnancy rate in the more general group of poor responders). This was strong enough

anecdotal evidence for physicians to be swayed away from the study. Interestingly, an abstract of a case-series was published regarding the use of letrozole to improve ovarian response in patients with a history of failed IVF (ie no pregnancy, no comment on E2 levels or oocyte numbers) who had used a minimum daily dose of 450 IU of gonadotropins (vague inclusion criteria). Again patients served as their own historical controls and the pregnancy rate after a letrozole IVF cycle approached 30%.(113) The migration to this number of 30% and failure to apply rigorous methodology likely is a result of selection bias. But, it is an excellent example of the predominant type of evidence used to determine treatment in this field.

### **3.3.3.3 Ethics**

The ethics of this trial could be a subject for debate. The clinical question is certainly ethical as there is equipoise as to the better treatment alternative. Both theoretical and clinical equipoise exist and we feel we are offering our patients a service by trying to answer this question. (119) We feel that given the emotional and financial burdens of disease we have a responsibility to determine in the most valid way what the better treatment is rather than continue with anecdotal evidence.

The patient, however, may feel differently. Couples facing infertility treatment have made huge financial, emotional and physical investments. For those couples who have already failed standard therapies and some of whom have traveled across the world to an internationally renowned institution and all of whom pay heavily out of pocket, they may not feel it is ethical to withhold a potentially beneficial treatment.

One of the obstacles in our first attempted PCOS RCT was the fact that patients had a good longterm prognosis to achieve pregnancy and the randomized treatments were already both standard of care. Patients had no “incentive” to participate. Given a choice, patients clearly prefer not to be a “guinea pig”. However, in a properly conducted trial, incentive should not be a coercing factor for subjects.

One of the criteria for subjects to make an informed consent about a RCT, or any medical therapy, is that they must be understand relevant information and be able to manipulate the information rationally.(120) The subjects who participated in this trial clearly experienced a degree of desperation. It is well established that infertility patients suffer from a difficulty to accept defeat of the situation and often do not attain closure. Many of the participants related to the research nurse that they were enrolling because they felt the PI had spent a considerable amount of time answering their questions and they felt content they would receive perhaps even a higher level of care. This in itself could be considered an unintentional form of coercion. Secondly, while this patient group tends to be very educated they may have a limited capacity to truly understand the equipoise nature of the question. Certainly, their decision to participate, despite all efforts on the part of the PI, may not be rational but rather out of desperation. The fact that almost, if not all, of the patients indicated a preference for the experimental protocol, some of whom burst into tears when they were assigned the standard protocol, attests to this. These patients clearly stated they would prefer to do the experimental protocol outside the study but were advised that there was equipoise regarding the benefit of the experimental protocol and it would therefore not be offered outside the study until or unless it had been proven beneficial. Nevertheless, some patients who did reapproach their primary physician after the debriefing session did not enroll because they were prescribed the experimental protocol outside the study.

#### **3.3.3.4 Summary**

In summary, the challenges associated with this RCT are considerable. Nevertheless, the question is important and the pursuit of a valid answer would be worthwhile. The feasibility study demonstrates that the study is achievable and many of the issues raised have reasonable potential solutions. My further research will involve pursuing this study, possibly in a multicentre scenario.

## Chapter 4 Discussion

### 4.1 Evidence based medicine

Perhaps the most valuable lesson learned during the implementation of this thesis was a true appreciation for evidenced based medicine and the challenges associated with obtaining evidence based medicine. There is considerable paucity of well-conducted controlled trials in the area of reproductive medicine. During my three fellowship in New York at the world-premier IVF institute, I observed a reluctance towards evidence based, particularly population based medicine. The attitudes were grounded in the belief that individualization of therapy and precise monitoring, management and attention to detail at every step of the IVF process is responsible for the unsurpassed success rates, particularly in extremely difficult cases. Some physicians even suggested that clinical research may be “unscientific” when compared to basic science research. Fortunately, this was not the attitude of the director which allowed this research to continue. The ambivalence regarding evidence based population medicine is not completely unfounded given the dogmatic and unindividualized, untailored approach of some clinical trials. Indeed, there will not be evidence for most of what we do as physicians and clinical intuition and experience clearly are invaluable in guiding medical decisions.

However, the lesson learned by the implementation of the steps of this thesis is that pursuit of evidenced-based medicine, where possible, is a valuable act. Flexibility and respect for circumstances at given trial centres in combination with perseverance and tenacity will aid in the attainment of knowledge. In application of clinical trial results, the importance of validity as well as generalizability is often reiterated.(121) However, where there is a paucity of any sort of information, valid information obtained through rigorous methodology, even though the population may not be generalizable, is a worthwhile start. Extrapolation to a similar population from carefully obtained evidence may be a better approach than no approach at all when there is no consensus about the utility of a treatment.

The process of systematic review greatly aided study design. Before subjecting patients to experimental therapies, or randomization of existing therapies, the trialist has a responsibility to qualitatively and quantitatively assimilate and summarize the available published and unpublished evidence in order to justify the trial. The trialist also owes this to the potential funding agencies. In addition, the systematic review process often helps redefine and narrow the study question.

#### **4.2 Challenges of Developing and Implementing Randomized Controlled Trial Protocols in Reproductive Medicine**

We have addressed these difficulties at many stages in this thesis. The following list is a summary of these points:

- 1) The question must be one of true clinical, not just theoretical, equipoise otherwise there is unlikely to be physician commitment.
- 2) The trial environment should have a culture which is amenable to the pursuit of evidence base medicine – the capitalistic nature of IVF and competing business interests of clinics in densely saturated areas may interfere with this.
- 3) A trial may be easier run in an environment where outcomes are less successful but the results may not be as informative. Highly prestigious institutes with international reputations based on success rates may have more to lose by participating in a trial that does not respect their management styles. Therefore adaptability to the trial centre is important.
- 4) Physician enthusiasm is often overestimated and wanes with time if results are not provided. Rapid gratification of results (eg time to pregnancy) is a characteristic of this group. Trialists must address this phenomenon.
- 5) The high direct costs to the patient associated with IVF make subjects less likely to participate in an RCT unless their prognosis is poor and the treatment empirical and unavailable elsewhere.
- 6) Peer-reviewed funding of studies in this area is very limited. However, funding might offset some of the reluctance of patients to participate, since the cost to them

is high, and facilitate implementation of consistent management protocols for multicentre trials.

- 7) Recruitment issues are significant for reasons mentioned earlier. Depending on the nature of the study a single center versus multicentre trial may be most appropriate. There is a tendency to skew enrollment to the worst prognosis subset. Specific efforts to enroll subjects in the range of the desired inclusion criteria may be necessary.
- 8) Challenges specific to the area of assisted reproduction include the following:
  - a) the high cost of unfunded therapy influences physician and patient participation;
  - b) business competition among clinics and patient demands often prevent clinics from participating in trials;
  - c) wide variation in success rates and treatment regimens between different clinics challenge the validity of multicentre trials;
  - d) physician reluctance to relinquish decision making to inflexible treatment schedules also prevents participation;
  - e) ascertainment of pregnancy as an outcome may be biased if biochemical measure or clinical measures such as missed menses are used. Presence of gestational sac on ultrasound is a hard outcome and avoids the previous problem but does not evaluate the most important outcome which is healthy infant deliveries. Absence of continuity once the patients are referred to obstetricians presents a potentially surmountable obstacle.

### **4.3 The Challenge of MCID**

The determination of minimal clinically important difference (MCID) is controversial and a potentially challenging step in clinical trial design. MCID has many definitions depending on the perspective one takes. From a physician perspective it is the smallest difference in outcome measures that is perceived as beneficial and would lead to a change in patient management. MCID may be determined from existing literature, consensus, or expert opinion.

In the first trial the MCID was 15%. In the second trial the MCID is 20%. The first study involved patients with a good prognosis to become pregnant and a reasonable regression to the mean as evidenced by a 30% ovulation rate in the placebo group in our meta-analysis. The intervention in the first group is “low technology”, not invasive, nor expensive. Therefore, in consultation with the clinician group at our center, it was decided that metformin should be at least 15% better than CC which has been standard therapy for 30 years in order to change their practice, particularly given their reluctance to accept the proposed mechanism of action of metformin. Attitudes toward metformin vary widely between clinical centers.

The second trial involved a group of patients with a much poorer prognosis (actual pregnancy rate in the control group 3.8%), undergoing invasive and expensive therapy. The regression to the mean is essentially negligible. If pregnancy is achieved, however, the miscarriage rate in this group is high (> 40%). Therefore, we proposed a higher MCID of 20%. This number was arrived at by consensus by 2 experts in the field. The number, however, was used for the theoretical trial planned in advance of the pilot study. Pilot studies often provide revised estimates of baseline rates and so impact the choice of MCIDs. In fact, the control rate was 3.8 % rather than the expected 10%. The choice of a 20% MCID was influenced by the fact that the expected pregnancy rate was 10% and a 20% increase would result in a 30% pregnancy rate, which is generally the minimal acceptable pregnancy rate in non poor responder patients with otherwise uncomplicated histories and good prognoses undergoing IVF.

One important result of the pilot study was the lower than expected control pregnancy rate of 3.8%. This will influence the determination of the MCID for the definitive study. We will need to survey physicians in this area regarding what would be the smallest difference that would be perceived as beneficial given that the control pregnancy rate is approximately 4%. (much lower than the expected 10% used in the previous determination of MCID). The revised acceptable pregnancy rate may now be lower (possibly around 10%). Since the original trial was designed, the American Society for Reproductive Medicine has published a policy statement on therapy in cases where the prognosis is

futile, very poor, or poor. Futile was defined as a success rate of less than 1%, very poor defined as less than 5%, and poor defined as less than 10%. When the expected success rate is less than 10%, treatment is generally discouraged by physicians but is often desired by patients and may be considered in individual cases.

#### **4.4 Recommendations**

There is an urgent need to evaluate, in a systematic and rigorous fashion, clinical outcomes of current and novel reproductive technologies. The challenges associated with such investigation are significant but not insurmountable. Collaboration between specialized centres, flexibility in clinical management strategies of research protocols, education of specialty physicians involved in this area, and carefully designed studies as a result of physician and epidemiologist collaboration will aid in these endeavors.

The next step will be to secure funding for a larger definitive trial of poor responders.



## References

1. Ron E, Lunenfeld B, Menczer J, Blumstein T, Katz L, Oelsner G, et al. Cancer incidence in a cohort of infertile women. *American Journal of Epidemiology*. 1987;125(5):780-90.
2. Collins J. Cost-effectiveness of in vitro fertilization. *Seminars in Reproductive Medicine*. 2001;19(3):279-89.
3. Jain T, Harlow BL, Hornstein MD. Insurance Coverage and Outcomes of in Vitro Fertilization. *N Engl J Med* 2002;347(9):661-666.
4. Heitman E. Social and ethical aspects of in vitro fertilization. 22-35, 1999 Winter.
5. Santiago-Delefosse M, Cahen F, Coeffin-Driol C. [The analysis of physicians' work: announcing the end of attempts at in vitro fertilization]. 293-305, 2003 Jul-Aug.
6. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *Jama*. 2000;283(15):2008-12.
7. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. [comment]. *Lancet*. 1999;354(9193):1896-900.
8. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. QUOROM Group. *British Journal of Surgery*. 2000;87(11):1448-54.
9. Barbieri RL. Polycystic ovarian disease. *Annual Review of Medicine*. 1991;42:199-204.
10. Dewailly D. Definition and significance of polycystic ovaries. *Baillieres Clinical Obstetrics & Gynaecology*. 1997;11(2):349-68.
11. Velazquez EM, Mendoza S, Hamer T, Sosa F, Glueck CJ. Metformin therapy in polycystic ovary syndrome reduces hyperinsulinemia, insulin resistance, hyperandrogenemia, and systolic blood pressure, while facilitating normal menses and pregnancy. *Metabolism: Clinical & Experimental*. 1994;43(5):647-54.

12. Moher D. CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. Consolidated Standards of Reporting Trials. *Jama*. 1998;279(18):1489-91.
13. Mitwally MF, Casper RF. Use of an aromatase inhibitor for induction of ovulation in patients with an inadequate response to clomiphene citrate. *Fertility & Sterility*. 2001;75(2):305-9.
14. Fisher SA, Reid RL, Van Vugt DA, Casper RF. A randomized double-blind comparison of the effects of clomiphene citrate and the aromatase inhibitor letrozole on ovulatory function in normal women. *Fertility & Sterility*. 2002;78(2):280-5.
15. Healy S, Sylvestre C, Tan SL, Biljan M. A comparison between superovulation with FSH and a combination of FSH and letrozole. *Fertility & Sterility*. 2002;76(3S):S55 - S56.
16. Franks S, Adams J, Mason H, Polson D. Ovulatory disorders in women with polycystic ovary syndrome. *Clinics in Obstetrics & Gynaecology*. 1985;12(3):605-32.
17. Nestler JE, Stovall D, Akhter N, Iuorno MJ, Jakubowicz DJ. Strategies for the use of insulin-sensitizing drugs to treat infertility in women with polycystic ovary syndrome.[comment]. *Fertility & Sterility*. 2002;77(2):209-15.
18. Homburg R. Should patients with polycystic ovarian syndrome be treated with metformin?: A note of cautious optimism. *Hum. Reprod*. 2002;17(4):853-856.
19. Stadtmauer LA, Wong BC, Oehninger S. Should patients with polycystic ovary syndrome be treated with metformin?: Benefits of insulin sensitizing drugs in polycystic ovary syndrome--beyond ovulation induction. *Hum. Reprod*. 2002;17(12):3016-3026.
20. Brinton LAP, Lamb EJMD, Moghissi KSMD, Scoccia BMD, Althuis MDP, Mabie JE, et al. Ovarian Cancer Risk After the Use of Ovulation-Stimulating Drugs. *Obstetrics & Gynecology* June 2004;103(6):1194-1203.
21. Rossing MA, Daling JR, Weiss NS, Moore DE, Self SG. Ovarian tumors in a cohort of infertile women. *New England Journal of Medicine*. 1994;331(12):771-6.
22. Hughes E, Collins J, Vandekerckhove P. Clomiphene citrate for ovulation induction in women with oligo-amenorrhoea. CD000056, 2000.

23. Asuncion M, Calvo RM, San Millan JL, Sancho J, Avila S, Escobar-Morreale HF. A Prospective Study of the Prevalence of the Polycystic Ovary Syndrome in Unselected Caucasian Women from Spain. *J Clin Endocrinol Metab* 2000;85(7):2434-2438.
24. Knochenhauer ES, Key TJ, Kahsar-Miller M, Waggoner W, Boots LR, Azziz R. Prevalence of the Polycystic Ovary Syndrome in Unselected Black and White Women of the Southeastern United States: A Prospective Study. *J Clin Endocrinol Metab* 1998;83(9):3078-3082.
25. Dunaif A. Insulin Resistance and the Polycystic Ovary Syndrome: Mechanism and Implications for Pathogenesis. *Endocr Rev* 1997;18(6):774-800.
26. Diamanti-Kandarakis E, Kouli C, Tsianateli T, Bergiele A. Therapeutic effects of metformin on insulin resistance and hyperandrogenism in polycystic ovary syndrome. *European Journal of Endocrinology* 1998;138(3):269-274.
27. Adams J, Polson DW, Franks S. Prevalence of polycystic ovaries in women with anovulation and idiopathic hirsutism. *British Medical Journal* 1986;293:355-359.
28. Hull MG. Epidemiology of infertility and polycystic ovarian disease: endocrinological and demographic studies. *Gynecologic Endocrinology* 1987;1:235-245.
29. Balen A, Michelmore K. What is polycystic ovary syndrome?: Are national views important? *Hum. Reprod.* 2002;17(9):2219-2227.
30. Dale P, Tanbor T, Vaaler S. Body weight, hypersinulinemia, and gonadotropin levels in the polycystic ovarian syndrome: Evidence of two distinct populations. *Fertility & Sterility.* 1992;58:487-91.
31. ESHRE, ASRM. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome\*1. *Fertility and Sterility* 2004;81(1):19-25.
32. Carmina E, Lobo RA. Polycystic ovaries in hirsute women with normal menses. *The American Journal of Medicine* 2001;111(8):602-606.
33. Legro RS, Driscoll D, Strauss JF, III, Fox J, Dunaif A. Evidence for a genetic basis for hyperandrogenemia in polycystic ovary syndrome. *PNAS* 1998;95(25):14956-14960.
34. Abbott DH, Dumesic DA, Franks S. Developmental origin of polycystic ovary syndrome - a hypothesis. *Journal of Endocrinology.* 2002;174(1):1-5.

35. Elting MW, Korsen TJM, Rekers-Mombarg LTM, Schoemaker J. Women with polycystic ovary syndrome gain regular menstrual cycles when ageing. *Hum. Reprod.* 2000;15(1):24-28.
36. Hughesdon PE. Morphology and morphogenesis of the Stein-Leventhal ovary and of so-called "hyperthecosis". *Obstet Gynecol Surv* 1982;37(2):59-77.
37. Elkind-Hirsch K, Valdes C, McConnel T. Androgen responses to acutely increased endogenous insulin levels in hyperandrogenic and normal cycling women. *Fertility & Sterility.* 1991;55:486.
38. Loverro G, Lorusso F, De Pagola, Nicolardi G, Mei L, Selvaggi L. Clinical and endocrinological effects of 6 months of metformin treatment in young hyperinsulinemic patients affected by polycystic ovarian syndrome. *Gynecological Endocrinology* 2002;16(3):217-24.
39. Mitkov M, Simenov S, Pehlivanov B. Effect of metformin treatment of polycystic ovary syndrome. *Endocrinologia* 2002;7(1):25-29.
40. Nagamani M, Dinh T, Kelder M. Hyperinsulinemia in hyperthecosis of the ovaries. *American Journal of Obstetrics and Gynecology* 1986;154:384.
41. Kim LH, Taylor AE, Barbieri RL. Insulin sensitizers and polycystic ovary syndrome: can a diabetes medication treat infertility? \*1. *Fertility and Sterility* 2000;73(6):1097-1098.
42. Judd HL, Barnes AB, Kliman B. Long-term effect of wedge resection on androgen production in a case of polycystic ovarian disease. *American Journal of Obstetrics & Gynecology.* 1971;110(8):1061-5.
43. Adelusì B, Aimakhu VE, Wright EA. Polycystic ovaries. Presentation and response to wedge resection. *International Journal of Gynaecology & Obstetrics.* 1976;14(3):232-6.
44. Greenblatt E, Casper RF. Endocrine changes after laparoscopic ovarian cautery in polycystic ovarian syndrome. *American Journal of Obstetrics & Gynecology.* 1987;156(2):279-85.
45. Farquhar C, Vandekerckhove P, Lilford R. Laparoscopic "drilling" by diathermy or laser for ovulation induction in anovulatory polycystic ovary syndrome.[update of

Cochrane Database Syst Rev. 2000;(2):CD001122; PMID: 10796746]. Cochrane Database of Systematic Reviews. 2001(4):CD001122.

46. Norman RJ, Davies MJ, Lord J, Moran LJ. The role of lifestyle modification in polycystic ovary syndrome. *Trends in Endocrinology & Metabolism*. 2002;13(6):251-7.
47. Huber-Bucholz MM, Carey DG, Norman RJ. Restoration of reproductive potential by lifestyle modification in obese polycystic ovary syndrome: role of insulin sensitivity and luteinizing hormone. *J Clin Endocrinol Metab* 1999;84:1470-4.
48. Pasquali R, Antenucci D, Casimirri F, Venturoli S, Paradisi R, Fabbri R, et al. Clinical and hormonal characteristics of obese amenorrheic hyperandrogenic women before and after weight loss. *J Clin Endocrinol Metab* 1989;68(1):173-179.
49. Clark A, Thornley B, Tomlinson L, Galletley C, Norman R. Weight loss in obese infertile women results in improvement in reproductive outcome for all forms of fertility treatment. *Hum. Reprod*. 1998;13(6):1502-1505.
50. Wysowski DE. Use of fertility drugs in the United States, 1979 through 1991. *Fertility & Sterility*. 1993;60:1096-98.
51. Gorbatsky G, Kase N, Speroff L. Ovulation and pregnancy rates with clomiphene citrate. *Obstet Gynecol* 1978;51(3):265-269.
52. Imani B, Eijkemans MJC, te Velde ER, Habbema JDF, Fauser BCJM. Predictors of Patients Remaining Anovulatory during Clomiphene Citrate Induction of Ovulation in Normogonadotropic Oligoamenorrheic Infertility. *J Clin Endocrinol Metab* 1998;83(7):2361-2365.
53. Imani B, Eijkemans MJC, te Velde ER, Habbema JDF, Fauser BCJM. Predictors of Chances to Conceive in Ovulatory Patients during Clomiphene Citrate Induction of Ovulation in Normogonadotropic Oligoamenorrheic Infertility. *J Clin Endocrinol Metab* 1999;84(5):1617-1622.
54. Kousta E, White DM, Franks S. Modern use of clomiphene citrate in induction of ovulation. *Human Reproduction Update* 1997;3:359.
55. Lunenfeld B, Pariente C, Dor J, Menashe Y, Seppala M, Mortman H, et al. Modern aspects of ovulation induction. *Annals of the New York Academy of Sciences* 1991;626:207- 216.

56. Eijkemans MJC, Imani B, Mulders AGMGJ, Habbema JDF, Fauser BCJM. High singleton live birth rate following classical ovulation induction in normogonadotrophic anovulatory infertility (WHO 2). *Hum. Reprod.* 2003;18(11):2357-2362.
57. Nugent D, Vandekerckhove P, Hughes E, Arnot M, Lilford R. Gonadotrophin therapy for ovulation induction in subfertility associated with polycystic ovary syndrome (Cochrane Review). In: *The Cochrane Library*. John Wiley & Sons, Ltd. ed: John Wiley & Sons, Ltd.; 2004.
58. Radwanska E, Rawlins RG, Tummon I, Maclin V, Binor Z, Dmowski WP. Successful use of gonadotropin-releasing hormone agonist leuprolide for in vitro fertilization in a patient with polycystic ovarian disease and infertility unresponsive to standard treatment. *Fertility & Sterility.* 1988;49(2):356-9.
59. Salat-Baroux J, Alvarez S, Antoine JM, Cornet D, Tibi C, Plachot M, et al. Results of IVF in the treatment of polycystic ovary disease. *Human Reproduction.* 1988;3(3):331-5.
60. Buyalos RP, Lee CT. Polycystic ovary syndrome: pathophysiology and outcome with in vitro fertilization.[see comment]. *Fertility & Sterility.* 1996;65(1):1-10.
61. Dale PO, Tanbo T, Abyholm T. In-vitro fertilization in infertile women with the polycystic ovarian syndrome. *Human Reproduction.* 1991;6(2):238-41.
62. Dor J, Shulman A, Levran D, Ben-Rafael Z, Rudak E, Mashiach S. The treatment of patients with polycystic ovarian syndrome by in-vitro fertilization and embryo transfer: a comparison of results with those of patients with tubal infertility. *Human Reproduction.* 1990;5(7):816-8.
63. Speroff L, Fritz MA. Polycystic ovarian syndrome. In: *Clinical Gynecologic Endocrinology and Infertility*. 7 ed. Philadelphia: Lippincott, Williams and Wilkins; 2004.
64. Velazquez E, Mendoza S, Hamer T, Sosa F, Glueck C. Metformin therapy in polycystic ovarian syndrome reduces hyperinsulinemia, insulin resistance, hyperandrogenemia, and systolic blood pressure, while facilitating normal menses and pregnancy. *Metabolism: Clinical and Experimental.* 1994;43(5):647 -54.
65. Glueck C, Wang P, Fontaine R, Tracy T, Sieve-Smith L. Metformin-induced resumption of normal menses in 39 of 43 (91%) previously amenorrheic women with polycystic ovarian syndrome. *Metabolism: Clinical and Experimental.* 1999;48(4):511-9.

66. Moghetti P, Castello R, Negri C, Tosi F, Perrone F, Caputo M, et al. Metformin Effects on Clinical Features, Endocrine and Metabolic Profiles, and Insulin Sensitivity in Polycystic Ovary Syndrome: A Randomized, Double- Blind, Placebo-Controlled 6-Month Trial, followed by Open, Long-Term Clinical Evaluation. *The Journal of Clinical Endocrinology and Metabolism* 2000;85(1):139-145.
67. Glueck CJ, Wang P, Kobayashi S, Phillips H, Sieve-Smith L. Metformin therapy throughout pregnancy reduces the development of gestational diabetes in women with polycystic ovary syndrome. *Fertility & Sterility*. 2002;77(3):520-5.
68. Jakubowicz DJ, Iuorno MJ, Jakubowicz S, Roberts KA, Nestler JE. Effects of Metformin on Early Pregnancy Loss in the Polycystic Ovary Syndrome. *J Clin Endocrinol Metab* 2002;87(2):524-529.
69. Kjotrod S, von During V, Sunde A, Carlsen SM. Metformin treatment before IVF/ICSI in PCOS women: a prospective, randomized, double-blind study. *Fertility and Sterility* 2004;81(Supplement 3):7-9.
70. Kjotrod SB, von During V, Carlsen SM. Metformin treatment before IVF/ICSI in women with polycystic ovary syndrome; a prospective, randomized, double blind study. *Hum. Reprod*. 2004;19(6):1315-1322.
71. Jadad AR, Moher D, Klassen TP. Guides for reading and interpreting systematic reviews: II. How did the authors find the studies and assess their quality? *Archives of Pediatrics & Adolescent Medicine*. 1998;152(8):812-7.
72. Berlin JA, UoPM-ABGS. Does blinding of readers affect the results of meta-analysis? *Lancet*. 1997;350:185-86.
73. Jadad AR, Moore A, Carroll D, Jenkinson C, Reynolds JM, Gavaghan D, et al. Assessing the Quality of Reports of Randomized Clinical Trials: Is blinding necessary? *Controlled Clinical Trials*. 1996;17:1-12.
74. Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine*:321-59, 1999 Feb 15.
75. Flemming R, Hopkinson ZE, Wallace M, Greer IA, Sattar N. Ovarian Function and Metabolic Factors in Women with Oligomenorrhea Treated with Metformin in a Randomized Double Blind Placebo-Controlled Trial. *The Journal of Clinical Endocrinology and Metabolism* 2002;87(2):569-574.

76. Kocak M, Caliskan E, Simsir C, Haberal A. Metformin improves ovulatory rates, cervical scores, and pregnancy rates in clomiphene citrate -resistant women with polycystic ovary syndrome. *Fertility and Sterility* 2002;77(1):101-6.
77. Nestler JE, Jakubowicz DJ, Evans WS, Pasquali R. Effects of Metformin on Spontaneous and Clomiphene Induced Ovulation in the Polycystic Ovary Syndrome. *The New England Journal of Medicine* 1998;338:1876- 80.
78. Ng EHY, Wat NMS, Ho PC. Effects of Metformin on ovulation rate, hormonal and metabolic profiles in women with clomiphene -resistant polycystic ovaries:a randomized, double-blind placebo-controlled trial. *Human Reproduction* 2001;16(8):1625-1631.
79. Singh I, Bedaiwy MA, Hatwal A, Kumar A, Agarwal A. Increased pregnancy rates with Metformin and clomiphene citrate in non-obese patients with polycystic ovary syndrome: prospective randomized study. *Fertility and Sterility* 2001;76(3 S1):S94.
80. Sturrock NDC, Lannon B, Fay TN. Metformin does not enhance ovulation induction in clomiphene resistant polycystic ovary syndrome in clinical practice. *British Journal of Pharmacology* 2001;53:469-473.
81. Vandermolen DT, Ratts VS, Evans WS, Stovall DW, Kauma SW, Nestler JE. Metformin increases the ovulatory rate and pregnancy rate form clomiphene citrate in patients with polycystic ovary syndrome who are resistant to clomiphene citrate alone. *Fertility and Sterility* 2001;75(2):310-5.
82. Clark HD, Wells GA, Huet C, McAlister FA, Salmi LR, Fergusson D, et al. Assessing the Reliability of Randomized Trials: Reliability of the Jadad Scale. *Controlled Clinical Trials*. 1999;20:448-452.
83. Kashyap S, Wells GA, Rosenwaks Z. Insulin-sensitizing agents as primary therapy for patients with polycystic ovarian syndrome. *Hum. Reprod.* 2004:deh440.
84. Lord JM, Flight IHK, Norman RJ. Metformin in polycystic ovary syndrome: systematic review and meta-analysis. *BMJ* 2003;327(7421):951-0.
85. Yarali H, Yildiz BO, Demiroglu A, Zeyneloglu HB, Yigit N, Bukulmez O, et al. Co-administration of metformin during rFSH treatment in patients with clomiphene citrate-resistant polycystic ovarian syndrome: a prospective randomized trial. *Hum. Reprod.* 2002;17(2):289-294.



86. Jakubowicz DJ, Seppala M, Jakubowicz S, Rodriguez-Armas O, Rivas-Santiago A, Koistinen H, et al. Insulin Reduction with Metformin Increases Luteal Phase Serum Glycodelin and Insulin-Like Growth Factor-Binding Protein 1 Concentrations and Enhances Uterine Vascularity and Blood Flow in the Polycystic Ovary Syndrome. *J Clin Endocrinol Metab* 2001;86(3):1126-1133.
87. Nestler JE, Jakubowicz DJ. Decreases in Ovarian Cytochrome P450c17{alpha} Activity and Serum Free Testosterone after Reduction of Insulin Secretion in Polycystic Ovary Syndrome. *N Engl J Med* 1996;335(9):617-623.
88. Malkawi HY, Qulban HS. The effect of metformin plus clomiphene citrate on ovulation and pregnancy rates in clomiphene -resistant women with polycystic ovary syndrome. *Saudi Medical Journal* 2002;23:663-6.
89. El-Biely MM, Habba M. The use of metformin to augment the induction of ovulation in obese infertile patients with polycystic ovary syndrome. *Middle Eastern Fertility Society Journal* 2001;6:43-9.
90. Mitwally MF, Casper RF. Aromatase inhibition improves ovarian response to follicle -stimulating hormone in poor responders. *Fertility & Sterility*. 2002;77(4):776-780.
91. Buzdar AU. A summary of second-line randomized studies of aromatase inhibitors. *Journal of Steroid Biochemistry & Molecular Biology*. 2001;79(1-5):109-14.
92. Mitwally MF, Casper RF. Pregnancy outcome after the use of an aromatase inhibitor for ovarian stimulation. *Fertility and Sterility* 2002;78(Supplement 1):S277-S278.
93. Biljan M, Tan SL, Tulandi T. Prospective randomized trial comparing the effects of 2.5 and 5 mg of letrozole (LE) on follicular development, endometrial thickness, and pregnancy rate in patients undergoing superovulation. *Fertility & Sterility*. 2002;76(3S):S55.
94. Pincus G, Enzmann E. Can mammalian eggs undergo normal development in vitro? *Proceedures of the National Academy of Science USA* 1934;20:121.
95. Thibault C, Dauzier L, Wintenberger S. Etude cytologique de la fecondation in vitro de l'oeuf de la lapine. *C R Soc Biol Paris* 1954;148:789.

96. Perone N. In vitro fertilization and embryo transfer- a historical perspective. *The Journal of Reproductive Medicine* 1994;39:695-700.
97. Damario MA, Davis O, Rosenwaks Z. The role of maternal age in the assisted reproductive technologies. *Reproductive Medicine Review* 1999; 7:1 41-60 1999;7:141-60.
98. Liccardi F, Liu H, Rosenwaks Z. Day 3 estradiol serum concentrations as prognosticators of ovarian stimulation response and pregnancy outcome in patients undergoing in vitro fertilization. *Fertil Steril* 1995;64:991-4.
99. Navot D, Rosenwaks Z, Margalioth E. Prognostic assessment of female fecundity. *Lancet*. *Lancet* 1987;1(645-7).
100. Dunson DB, Colombo B, Baird DD. Changes with age in the level and duration of fertility in the menstrual cycle. *Human Reproduction*. 2002;17(5):1399-403.
101. Tietze C. Reproductive span and rate of reproduction among Hutterite women. *Fertil Steril* 1957;8:89.
102. Akman MA, Erden HF, Tosun SB, Bayazit N, Aksoy E, Bahceci M. Comparison of agonistic flare-up-protocol and antagonistic multiple dose protocol in ovarian stimulation of poor responders: results of a prospective randomized trial. *Human Reproduction*. 2001;16(5):868-70.
103. Akman MA, Erden HF, Tosun SB, Bayazit N, Aksoy E, Bahceci M. Addition of GnRH antagonist in cycles of poor responders undergoing IVF. *Human Reproduction*. 2000;15(10):2145-7.
104. Dirnfeld M, Fruchter O, Yshai D, Lissak A, Ahdut A, Abramovici H. Cessation of gonadotropin-releasing hormone analogue (GnRH-a) upon down-regulation versus conventional long GnRH-a protocol in poor responders undergoing in vitro fertilization. *Fertility & Sterility*. 1999;72(3):406-11.
105. Rombauts L, Suikkari AM, MacLachlan V, Trounson AO, Healy DL. Recruitment of follicles by recombinant human follicle-stimulating hormone commencing in the luteal phase of the ovarian cycle. *Fertility & Sterility*. 1998;69(4):665-9.
106. Bergh C, Carlstrom K, Selleskog U, Hillensjo T. Effect of growth hormone on follicular fluid androgen levels in patients treated with gonadotropins before in vitro fertilization. *European Journal of Endocrinology*. 1996;134(2):190-6.

107. Dor J, Bider D, Shulman A, Levron JL, Shine S, Mashiach S, et al. Effects of gonadotrophin-releasing hormone agonists on human ovarian steroid secretion in vivo and in vitro-results of a prospective, randomized in-vitro fertilization study. *Human Reproduction*. 2000;15(6):1225-30.
108. Bergh C, Hillensjo T, Wikland M, Nilsson L, Borg G, Hamberger L. Adjuvant growth hormone treatment during in vitro fertilization: a randomized, placebo-controlled study. *Fertility & Sterility*. 1994;62(1):113-20.
109. Hughes SM, Huang ZH, Morris ID, Matson PL, Buck P, Lieberman BA. A double-blind cross-over controlled study to evaluate the effect of human biosynthetic growth hormone on ovarian stimulation in previous poor responders to in-vitro fertilization. *Human Reproduction*. 1994;9(1):13-8.
110. Lanzendorf SE, Nehchiri F, Mayer JF, Oehninger S, Muasher SJ. A prospective, randomized, double-blind study for the evaluation of assisted hatching in patients with advanced maternal age. *Human Reproduction*. 1998;13(2):409-13.
111. Juni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study.[comment]. *International Journal of Epidemiology*. 2002;31(1):115-23.
112. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, et al. What contributions do languages other than English make on the results of meta-analyses? *Journal of Clinical Epidemiology*. 2000;53(9):964-72.
113. Quintero R, Giudice LC, Westphal L. Efficacy of aromatase inhibitors in poor responders undergoing in vitro fertilization. *Fertility and Sterility* 2004;81(Supplement 3):28.
114. Mitwally MF, Casper RF. Single dose administration of the aromatase inhibitor, letrozole: a simple and convenient effective method of ovulation induction. *Fertility and Sterility* 2001;76(3, Supplement 1):S94-S95.
115. Friedman LM, Furberg CD, DeMets DL. *Fundamentals of Clinical Trials*. 3rd ed. New York: Springer - Verlag New York, Inc.; 1998.
116. Chalmers TC, Celano P, Sacks HS, Smith H. Bias in treatment assignment in controlled clinical trials. *The New England Journal of Medicine* 1983;309:1358-61.

117. Schultz KF. Subverting randomization in controlled trials. *JAMA* 1995;272(8):1456-1458.
118. Assisted reproductive technology in the United States: 2000 results generated from the American Society for Reproductive Medicine/Society for Assisted Reproductive Technology Registry. *Fertility and Sterility* 2004;81(5):1207-1220.
119. Freedman B. Equipoise and the ethics of clinical research. *The New England Journal of Medicine* 1987;317:141-5.
120. Applebaum PS, Grisso T. Assessing patients capacities to consent to treatment. *The New England Journal of Medicine* 1988;319(25):1635-8.
121. Guyatt GH, Sinclair J, Cook DJ, Glasziou P, for the Evidence-Based Medicine Working Group and the Cochrane Applicability Methods Working Group. Users' Guides to the Medical Literature: XVI. How to Use a Treatment Recommendation. *JAMA* 1999;281(19):1836-1843.

## Appendix A

### 1. Jadad Quality Scale for assessment of RCTS

#### Scoring the items:

Either give a score of 1 point for each "yes" or 0 points for each "no." There are no in-between marks.

**Give 1 additional point if:** For question 1, the method to generate the sequence of randomization was described and it was appropriate (table of random numbers, computer generated, etc.)

**and/or:** If for question 2 the method of double blinding was described and it was appropriate (identical placebo, active placebo, dummy, etc.)

**Deduct 1 point if:** For question 1, the method to generate the sequence of randomization was described and it was inappropriate (patients were allocated alternately, or according to date of birth, hospital number, etc.)

**and/or:** For question 2, the study was described as double blind but the method of blinding was inappropriate (e.g., comparison of tablet vs. injection with no double dummy)

#### Guidelines for Assessment

##### 1. Randomization

A method to generate the sequence of randomization will be regarded as appropriate if it allowed each study participant to have the same chance of receiving each intervention and the investigators could not predict which treatment was next. Methods of allocation using date of birth, date of admission, hospital numbers, or alternation should be not regarded as appropriate.

##### 2. Double blinding

A study must be regarded as double blind if the word "double blind" is used. The method will be regarded as appropriate if it is stated that neither the person doing the assessments nor the study participant could identify the intervention being assessed, or if in the absence of such a statement the use of active placebos, identical placebos, or dummies is mentioned.

##### 3. Withdrawals and dropouts

Participants who were included in the study but did not complete the observation period or who were not included in the analysis must be described. The number and the reasons for withdrawal in each group must be stated. If there were no withdrawals, it should be stated in the article. If there is no statement on withdrawals, this item must be given no points.

{Jadad, 1996 #101}



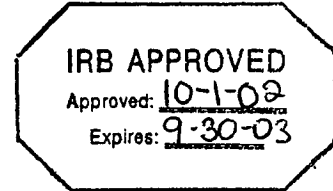
**Appendix B2**  
**PCOS RCT Patient Consent**

Date (Mo., Dy., Yr)	Location	Service
Age	Doctor	

If No Plate, Print Name, Sex, And History No.

**THE NEW YORK PRESBYTERIAN HOSPITAL-WEILL MEDICAL COLLEGE OF CORNELL UNIVERSITY**

**Consent Form for Clinical Investigation**



**Project Title:** Metformin versus Clomid for Ovulation Induction in PCOS – A  
Randomized Controlled Trial

**Subject:** \_\_\_\_\_ **Research Project #** \_\_\_\_\_

1. You are invited to participate in a research study of clomiphene citrate versus metformin for ovulation induction. Physicians of The New York Presbyterian Hospital-Weill Medical College of Cornell University hope to learn if metformin is as good or better than clomiphene citrate for ovulation induction. You were selected as a possible participant in this study because you have a diagnosis of polycystic ovarian disease and your physician is considering treating you with either clomiphene citrate or metformin.

The treatments described are the standard, first line therapies used for patients with your histories. Your clinical management will not be different than if you were not part of this study. Although both drugs are used as first line therapies in common practice, metformin is “off-label” which means the FDA when they approved it did so for its use in patients with diabetes. While we use it commonly for induction of ovulation in patients with your history, the FDA has not revised the indications for it.

By the end of therapy, the cost of either treatment should be similar.

1. If you decide and are eligible (by inclusion/exclusion criteria) to participate, you will be randomized (“like the flip of a coin”) you to either metformin or clomiphene citrate. All patients will also be referred for nutritional counseling. The minimum duration of time for the study will be 1 month but the maximum duration of time will be 6 months. This amount of time and the treatments undertaken do not differ in any way from how you would be clinically managed if you were not a participant in this study. This study will hope to answer if these treatments are equally effective or if one is better than the other. Patients who receive metformin will be treated for a minimum of 1 month and a maximum of 6 months. They will take the medication every day. Patients who receive clomiphene citrate will be on it for a minimum of 1 month and a maximum of 5 months. They will take it for five days of the cycle each month, if they do not become pregnant. All patients will be monitored for ovulation with ultrasound and blood tests as would be normally done if they were not participants in this study.



2. The costs for treatment will be billed to you or to your insurance as they are consistent with routine clinical practice. Subjects do not receive any compensation for their participation.

Participation requires the following:

- a. A special visit with the study team for physical examination and blood tests. The amount of blood that will be drawn will equal approximately 3 tablespoons at this first visit. You will also be assessed to ensure that you meet the trial requirements (inclusion/exclusion criteria)
  - b. IF your initial bloodwork is normal, we will bring you back for a second visit to obtain informed consent, obtain pre-treatment bloods tests (Approximately 1 tablespoon), randomize you to a treatment, and give you a prescription for the treatment. The cost of medications is the responsibility of the patient or the patient's insurance.
  - c. All participants must have a partner with normal semen testing. This test will have been done as part of your initial work-up with your physician as per usual infertility testing. All participants will be advised to have intercourse 2 –3 times a week. All patients will also be referred for nutritional counseling. The nutritionist to whom we refer accepts most insurances. If cost is an issue, we can arrange alternative counseling.
  - d. Patients who receive clomiphene treatment will need to return for blood tests (approximately 1 tablespoon) on days 5, 7, and 10 after the last pill to determine ovulation. They will also have ultrasound monitoring at these times. IF ovulation occurs, monitoring of menses and blood tests (1 tablespoon monthly) will be done for 3 months to determine if a subject becomes pregnant. If after 3 months of ovulation, no pregnancy occurs, treatment will be considered a failure. Alternative options will be considered. If ovulation does not occur the first month, the dose will be stepped up and similar monitoring to the above will occur. If ovulation does not occur at the 2<sup>nd</sup> dose, the dose will be stepped up one final time and monitoring repeated as above. IF Ovulation does not occur at the 3<sup>rd</sup> dose, treatment will be considered a failure and alternative options considered. This protocol is standard procedure for patients receiving clomiphene. The cost of monitoring will be consistent with routine clinical screening.
  - e. Patients who receive metformin will receive 500mg /day for one week, then increase the dose to 500mg twice a day for the second week and finally increase the dose to 850 mg twice a day in the third week. This is the dose of medication you will take for the remainder of the treatment. You will return every 2 weeks for blood tests (1 tablespoon) and ultrasound to monitor for ovulation until ovulation occurs or 2 months passes. If ovulation does not occur by two months, treatment will be considered a failure and alternative options considered. If ovulation occurs before 2 months, and once ovulation occurs patients will be monitored for pregnancy by monitoring of menses and pregnancy tests. (Blood : approximately 1 tablespoon monthly)
- c) Your participation in the project involves the following risks:
1. patients who take clomiphene citrate have a 10% risk of having twins. They have a very minimal risk (less than 1%) of having enlarged ovaries (ovarian hyperstimulation)

**THE NEW YORK PRESBYTERIAN HOSPITAL**

(continued)

(Use other side if necessary)

2. **with metformin fewer than 1 in 33 000 women an illness called "lactic acidosis" if you have kidney or liver disease. Since this disorder can be fatal, you will not be eligible for the study if you have a history of liver or kidney disease or if your liver or kidney blood tests are abnormal.**
  3. **patients who receive metformin have a 20% incidence of diarrhea if they do not comply with dietary recommendations**
  4. **patients who receive clomiphene may experience the following: 10% hot flushes; 2% breast tenderness; Abdominal distension/bloating 5.5%; nausea and vomiting 2.2%, visual symptoms 1.5%; headache 1.3%; dryness or loss of hair 0.3%.**
  5. **Alcohol should not be consumed during treatment for two reasons: it is not advisable to consume alcohol while trying to conceive and alcohol can negatively impact the metabolism of metformin.**
- d) We cannot and do not guarantee that you will receive any benefits from this study. Pregnancy is the goal of treatment as it would be with or without the study. The general benefits to society are that the question as to whether or not metformin is as effective as clomiphene will hopefully be answered.
  - e) Any information obtained during this study and identified with you will remain confidential and will be disclosed only with your permission. The exception to this is that the Institutional Review Board (IRB) and the Office for Human Research Protection (OHRP) may have access to the files.
  - f) Your decision whether or not to participate will not prejudice your future relations with The New York Presbyterian Hospital-Weill Medical College of Cornell University. If you decide to participate, you are free to discontinue participation at any time.
  - g) IF you decide not to participate, the usual treatment alternatives are still available to you but not in a randomized fashion. These alternatives still include therapy with metformin or clomiphene citrate or no treatment.
  - h) Reasons that we might discontinue you from the study include the following: IF you develop concurrent illnesses, particularly liver or kidney illnesses and you are receiving metformin, we will have to discontinue from the study. IF you develop side effects of treatment that are persistent or worrisome, we will have to discontinue you from the study. IF, for other reasons, you are unable to comply with the rules of the study, we will have to discontinue you from the study.
  - i) In accordance with Federal regulations, we are obligated to inform you about the Medical Center's policy in the event physical injury occurs. If, as a result of your participation, you experience physical injury from known or unknown risks of the research procedures as described, immediate medical care and treatment, including hospitalization, if necessary, will be available. No monetary compensation, however, is available and you will be responsible for the costs of such medical treatment, either directly or through your medical insurance and/or

**Appendix C1**  
**Published PCOS Meta-analysis: The Use of Insulin-Sensitizing Agents as Primary  
Therapy for Polycystic Ovarian Syndrome**

investigating metformin, CC, and metformin plus CC. Thirteen centres are involved in the recruitment of 768 patients with oligomenorrhoea, elevated testosterone, normal semen parameters and regular intercourse (2–3 times/week) who desire pregnancy.

## Materials and methods

One reviewer has completed all stages. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) and Quality Of Reporting Of Meta-analyses (QUOROM) guidelines were followed (Moher *et al.*, 1999; Stroup *et al.*, 2000).

### Search strategy

We used an Ovid vendor and Polaris interface to conduct a computerized literature search of the following seven bibliographic databases: Medline, Premedline, Current Contents, Biological Abstracts, Cochrane Controlled Trials Register, Cochrane Database of Systematic Reviews, and EMBASE. The databases were searched for the last 25 years. MeSH headings and textwords, for databases without MeSH headings, were used. Adjacency operators and truncation were used. A preliminary search was conducted to maximize potential key words. We did not apply a study filter and the search was not limited by language or year of publication (Moher *et al.*, 1996, 2000; Juni *et al.*, 2002). The search was run every 3 weeks between August, 2002 and September, 2003 to identify new articles.

A hand search of the following journals was also completed: *Fertility and Sterility*, *Human Reproduction*, *New England Journal of Medicine*, *Journal of Clinical Endocrinology*, and *The Lancet*. Online indexing facilitated this process.

To avoid publication bias, we attempted to search for 'grey literature' (McAuley *et al.*, 2000). Two content experts were contacted. No completed alternative studies were identified but an ongoing RCT was identified. Results are not available (private communication, Reproductive Medicine Network, <http://rmn.dcri.duke.edu/>).

Ten years of conference proceedings for the American Society of Reproductive Medicine were also searched. The proceedings for the Canadian Fertility and Andrology Society were searched from 1998 onward. One abstract for a RCT was found (Singh *et al.*, 2001).

The Cochrane Database was also searched for relevant papers. One protocol for a systematic review on the same topic was identified, but no completed review was found (Flight, 2002). The objective of the protocol was to assess multiple drugs of the same family and their effects on PCOS symptoms, sequelae (glucose intolerance, hypertension, cardiovascular disease), and adverse side-effects. This paper has, however, been published subsequently in both the Cochrane Database and the *British Medical Journal* and is considered in depth and compared to our review in the discussion (Lord *et al.*, 2003).

Titles and abstracts were screened and articles retrieved if they passed the relevance filter or if there was uncertainty as to whether or not they were relevant. Bibliographies of review articles, systematic reviews and retrieved studies were also searched for candidate articles (Jadad *et al.*, 1998). Retrieved articles were then reviewed for inclusion/exclusion criteria. Those articles that met the criteria were then kept for critical appraisal and data collection.

Reviewers were not blinded at any point to the authors or sources of publication as the evidence for such blinding is weak and the reviewers were likely to be previously familiar with some of the literature (Berlin, 1997).

## Inclusion/exclusion criteria

### Randomized controlled trials

**Study population.** The study population consists of women with primary or secondary infertility, between the ages of 18 and 40 years, who have been diagnosed with PCOS by the following characteristics: chronic oligo-ovulation (menstrual cycles less frequent than every 35 days or fewer than six periods per year); infertility; and one or more of the following characteristics: chronic hyperandrogenism (biochemical elevated testosterone, dehydroepiandrosterone, or androstenedione levels or clinical hirsutism or acne); increased LH/FSH ratio > 2.5; or ultrasound criteria of PCOS. The study and control groups should have no other infertility diagnosis and preferably would have a documented normal semen analysis.

**Intervention:** We looked for the following comparisons: Metformin versus placebo; metformin versus CC; metformin plus CC versus placebo plus CC.

**Outcome:** Two outcomes are assessed: ovulation as determined by serum progesterone level, and pregnancy as determined by urinary or serum  $\beta$ hCG.

### Cohort studies

Similar inclusion/exclusion criteria were applied but no cohort studies were found that met the criteria.

For studies that resulted in multiple publications, only the most recent or most complete publication was used. Relevant data were collected onto a pre-formed, standard data extraction sheet. The following were recorded: study characteristics (source, language, year, and design); subject characteristics (definition and selection of controls, study subjects, available information on confounders); intervention/exposure information (drug doses and duration); and outcome assessment (methods of ascertainment of exposure or outcome, and time to assessment of outcome).

The Jadad scale was used for quality assessment of RCT (Jadad *et al.*, 1996). Quality assessment involves evaluation of patient selection, assessment of exposure or outcomes, administration of interventions, and controls for confounding factors.

### Quantitative data synthesis

Where appropriate, Rev-man 4.1 and Metaview 4.0 have been used to analyse data. A relative risk (RR) estimate, with confidence intervals (CI), was extracted from RCT. We reported the fixed effects model since results did not differ from a random effects model. A  $\chi^2$ -test was done to determine the significance of the association. Heterogeneity is determined by the Cochran *Q*-test. These results are illustrated graphically in the form of a Forrest plot. Funnel plots have been constructed to represent the likelihood of publication bias. The precision of the search is determined by the formula of Normand.

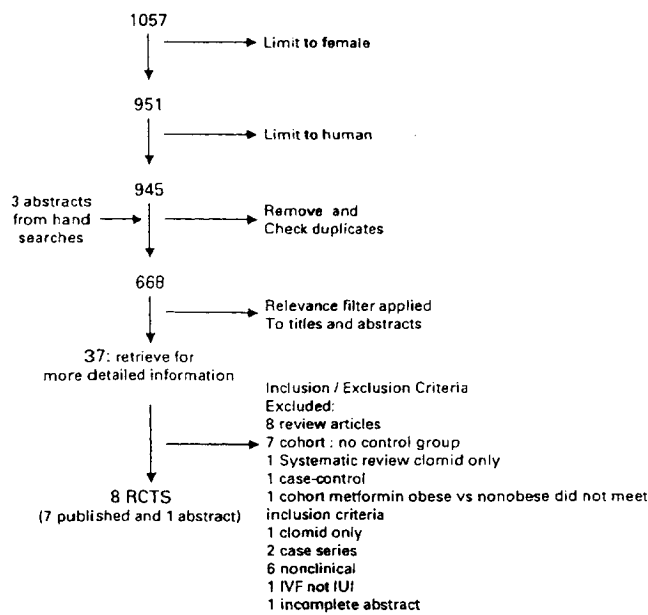
Where appropriate, sensitivity analysis and subgroup analysis were done to determine the significance of contributing factors to the overall results.

## Results

Overall precision of the search strategy was 0.76% (8/1057). Recall was 88% (7/8) (Figure 1).

### Language of publication

Although we did not limit the language of publication, all retrieved articles were in English (Moher *et al.*, 2000; Juni *et al.*, 2002).



**Figure 1.** A flow diagram showing the number of publications identified in the literature search. RCTS = randomized controlled trials. Generated from Meta-view 4.0.

#### Year of publication

All included RCT were published between 1998 and 2002 (Nestler *et al.*, 1998; Moghetti *et al.*, 2000; Ng *et al.*, 2001; Singh *et al.*, 2001; Sturrock *et al.*, 2001; Vandermolen *et al.*, 2001; Fleming *et al.*, 2002; Kocak *et al.*, 2002).

#### Quality of publication

The Jadad quality scale for RCTs was applied to all included studies and is shown in Table I (Clark *et al.*, 1999).

No RCTs were excluded for multiple publication. Studies were assessed for comparability of study population by age, body mass index, fertility, and method of diagnosis of PCOS. We also reviewed investigation for potential confounders, compliance, and contamination or cross-over. None of the studies commented on contamination or cross-over; only one study mentioned compliance but that study did not describe the methods used to assess compliance (Vandermolen *et al.*, 2001). One study described drop-outs but did not discuss compliance methods or if compliance was measured in continuing subjects (Fleming *et al.*, 2002).

All studies, except Singh *et al.* (2001), investigated metformin versus placebo. All studies except three used pregnancy as a primary outcome but only one abstract assessed pregnancy as a primary outcome (Nestler *et al.*, 1998; Moghetti *et al.*, 2000; Ng *et al.*, 2001). All studies except three had infertile, PCOS women as their study population (Nestler *et al.*, 1998; Moghetti *et al.*, 2000; Fleming *et al.*, 2002). Several studies tried to compare metformin plus CC to CC alone or with placebo (Nestler *et al.*, 1998; Ng *et al.*, 2001; Vandermolen *et al.*, 2001; Kocak *et al.*, 2002). However, all of these studies did so in a sequential manner in those patients who failed to conceive or ovulate with metformin. The duration of intervention and follow-up varies from 1 to 6 months. Table II illustrates the characteristics of these studies. Table III outlines objectives for each study.

Metformin induced ovulation in 47% of patients. Figures 2 and 3 indicate that metformin is 50% better than placebo for ovulation induction in infertile PCOS patients (RR 1.50) but this benefit is not necessarily improved with longer duration of therapy (> 3 months, RR 1.37). Metformin is also of benefit in non-infertile PCOS patients (i.e. patients with PCOS who were not complaining of infertility) for cycle regulation compared to placebo (Figure 3; RR 1.45 CI 1.11, 1.90). To date, metformin is not of confirmed benefit versus placebo for achievement of pregnancy (Figure 2; RR 1.07 CI 0.20, 5.74). This may be due to the fact that most studies used ovulation as their primary outcome, those studies that assessed pregnancy had a short follow-up time to pregnancy, and most studies had small sample sizes.

The analysis of metformin plus CC versus CC alone or with placebo is not properly represented here since the study design for this component is not a true RCT or cross-over design but rather a sequential study design. The nature of infertility therapy makes cross-over trials an inappropriate design. In order to conduct a cross-over trial, the two therapies being compared must be able to be 'washed out' within a short window of time between arms, the outcome must also be completely reversible within this window so that the patient returns to baseline, and the chance of successful outcome should not alter with time such that one therapy has a disadvantage over the other by being administered in the later arm or that the chance of success varies with time. Nevertheless, we did do a quantitative analysis on the data supplied to provide some assessment of the above

**Table I.** Quality assessment by Jadad scale

Study	Randomized?	Randomization method described*	Double-blind?	Double-blind method described*	Withdrawals and drop-outs described?	Score
Kocak <i>et al.</i> (2002) <sup>b</sup>	+1	-1	+1	+1	+1	3
Fleming <i>et al.</i> (2002) <sup>b</sup>	+1	+1	+1	+1	+1	5
Sturrock <i>et al.</i> (2001) <sup>b</sup>	+1		+1			2
Vandermolen <i>et al.</i> (2001) <sup>b</sup>	+1	+1	+1	+1	+1	5
Ng <i>et al.</i> (2001) <sup>b</sup>	+1	+1	+1	+1		4
Singh <i>et al.</i> (2001)	+1					1
Moghetti <i>et al.</i> (2000) <sup>b</sup>	+1		+1	+1		3
Nestler <i>et al.</i> (1998) <sup>b</sup>	+1					1

\*Appropriate (+1), inappropriate (-1).

<sup>b</sup>Studies included in quantitative analysis.

Table II. Characteristics of randomized controlled trials

Study and intervention	No. of subjects: a. Study; b. Control	Ascertainment of outcome: A. Ovulation B. Pregnancy	Location	Intervention	Method of randomization	Outcome	Sources of bias
Kocak <i>et al.</i> (2002)	a. 23 b. 23	A, B	Turkey	cycle 1: metformin versus placebo; cycle 2: metformin + CC versus placebo + CC But hCG at US > 18mm 14 weeks metformin versus placebo	sequential odd numbers metformin; even numbers placebo computer-generated blocks of 4 sample size $2n = 310$ intention to treat	ovulation n/a placebo versus metformin; pregnancy: metformin 1/23 versus placebo 0/23 hCG given in 2nd arm ovulation: metformin 37/45 versus placebo 30/47	no comment no comment
Fleming <i>et al.</i> (2002)	a. 45 b. 47	A, B	UK	metformin versus placebo × 12 weeks metformin + CC versus placebo + CC × 1 month	double-blind placebo intention to treat computer-generated blocks of six intention to treat double-blind placebo	ovulation: metformin 0/12 versus placebo 1/14 pregnancy: metformin + CC 3/12 versus placebo + CC 2/14	no comment
Sturrock <i>et al.</i> (2001)	a. 12 b. 14	A, B	UK	metformin versus placebo × 7 weeks metformin + CC versus placebo + CC × max 6 cycles	computer-generated blocks of six intention to treat double-blind placebo	ovulation: metformin 1/12 versus placebo 1/15 metformin + CC 8/11 versus placebo + CC 3/14 pregnancy: metformin 0 versus placebo 0 metformin + CC 6/11 versus placebo + CC 1/14	no comment
Vandermolen <i>et al.</i> (2001)	a. 11 b. 15	A	USA	metformin versus placebo × 3 months metformin + CC versus placebo + CC × 1 cycle	computer-generated double-blind placebo sample size $2n = 16$ based on Nestler's data	ovulation: metformin 3/9 versus placebo 3/9 metformin + CC 2/4 versus placebo + CC 4/4	no comment
Ng <i>et al.</i> (2001)	a. 9 b. 9	B	China	metformin + CC versus placebo × 4 months	method not described	pregnancy: met 1/9 versus placebo 2/9 metformin + CC 2/4 versus placebo + CC 4/4	no comment
Singh <i>et al.</i> (2001)	a. 53 b. 47	B	India	6 months metformin versus placebo metformin versus placebo × 5 weeks, then metformin + CC versus placebo + CC	method not described	pregnancy: metformin + CC 31/53 versus cc 4/47	no comment
Moghetti <i>et al.</i> (2000)	a. 11 b. 11	A	Italy		method not described double-blind placebo	ovulation: metformin 5/11 versus placebo 0/11 improved menses	no comment
Nestler <i>et al.</i> (1998)	a. 35 b. 26	A Progesterone > 8 mg/ml	USA		method not described	ovulation: metformin 12/35 versus placebo 1/26 placebo + CC 2/25 versus metformin + CC 19/21	no comment

Table II. Characteristics of randomized controlled trials (continued)

Study	Study group: a. PCOS b. Infertile PCOS	1. Primary 2. Secondary a. Ovulation b. pregnancy	Criteria for diagnosis of PCOS	Confounders: a. Duration infertility b. Previous treatment c. Age d. Body mass index e. Hyperandrogenaemia f. Hyperinsulinaemia	Compliance assessed	Length of follow-up (no. of cycles)	Assessment for tubal or male factor	Assessment for contamination/ cross-over
Kocak <i>et al.</i> (2002)	B	A, B	oligomenorrhoea hyperandrogenemia multiple subcapsular follicles on ultrasound	b, c, d, e, f	no comment	cycle 1: metformin versus placebo cycle 2: metformin CC > versus placebo + CC but hCG at CC > 18 14 weeks metformin versus placebo	tubal and male factor ruled out	no comment
Fleming <i>et al.</i> (2002)	A, B* age < 35 years	A, B*	oligo/amenorrhoea Adams <i>et al.</i> ultrasound criteria	c, d, e, f	drop-outs reported but compliance assessment not described SS calculation no comment SS calculation	metformin versus placebo X	not assessed/ pits not desiring pregnancy advised to use barrier contraception no comment	no comment
Sturrock <i>et al.</i> (2001)	B. 18-40	A, B	oligomenorrhoeic	b Assess insulin, androgens but no comparison c, d, e, f	no comment SS calculation	6 months metformin versus placebo	no comment	no comment
Vandermolen <i>et al.</i> (2001)	B age 18-35 years	A, B	oligoovulation hyperandrogenaemia	a, b, c, d, e, f	1 patient excluded because 'non- compliance' compliance measures not described no comment	metformin versus placebo X 7 weeks metformin + CC versus placebo + CC X max six cycles	tubal & male factor ruled out	no comment
Ng <i>et al.</i> (2001)	B. age < 40 years	A	irregular cycles and anovulation' by midluteal progesterone CC resistant	a, b, c, d, e, f	no comment	metformin versus placebo X 3 months metformin + CC versus placebo + CC X 1 cycle metformin + CC versus CC X 4 months	tubal and male factor ruled out	no comment
Singh <i>et al.</i> (2001)	B	B	oligo/amenorrhoea 'ultrasound' LH/FSH > 2	c	no comment	metformin versus placebo X 6 months metformin versus placebo	no comment	no comment
Moggetti <i>et al.</i> (2000)	A. age 18-35 years	A	oligomenorrhoea hyperandrogenaemia normal GTT BMI > 28 kg/m <sup>2</sup>	c BMI > placebo group androgen > metformin group	no comment	metformin + CC versus placebo + CC X 5 weeks then metformin + CC versus placebo + CC	not assessed	no comment
Nestler <i>et al.</i> (1998)	A	A	oligomenorrhoea hyperandrogenaemia ultrasound normal GTT	b. 31 or 61 patients previously received CC c, d, e, f	no comment	metformin versus placebo X 5 weeks. then metformin + CC versus placebo + CC	abstinence/barrier contraception advised	no comment

CC = cloniphene citrate; GTT = glucose tolerance test; BMI = body mass index.

**Table III.** Study objectives of randomized controlled trials

Study	Stated objectives
Kocak <i>et al.</i> (2002)	To evaluate whether metformin administration improves ovarian response to CC, hyperandrogenism, insulin resistance, and cervical score in women with CC resistant PCOS
Fleming <i>et al.</i> (2002)	To investigate the effects of metformin on detailed ovarian function in women with oligomenorrhoea and PCOS who were treated with a randomized double-blind placebo-controlled trial of 16 weeks. Also measured changes in anthropometry, glycaemic indices, and lipid profiles
Sturrock <i>et al.</i> (2001)	To study the effects of metformin in a cohort of CC-resistant infertile females attending an infertility clinic to see whether in clinical practice pretreatment enhances ovulation and pregnancy rates
Vandermolen <i>et al.</i> (2001)	To determine whether metformin treatment increases the ovulation and pregnancy rates in response to CC in women who are resistant to CC alone
Ng <i>et al.</i> (2001)	To determine (i) the ovulation rate and (ii) the changes in hormonal metabolic profiles in women with CC-resistant PCO after taking metformin for 3 months
Moggetti <i>et al.</i> (2000)	To assess the effects on menstrual abnormalities of a 6 month course of metformin in a group of 23 subjects with PCOS and normal glucose tolerance
Nestler <i>et al.</i> (1998)	To determine whether reducing hyperinsulinaemia with metformin would increase the ovulatory response to CC in obese women with PCOS

CC = clomiphene citrate; PCOS = polycystic ovarian syndrome.

interventions for ovulation and pregnancy (Figure 4). Given the stated limitations of such a study design, we found that metformin plus CC may be superior to CC alone or with placebo with regards to ovulation (RR 3.04, CI 1.77, 5.24) and pregnancy (RR 3.65, CI 1.11, 11.99).

Funnel plots for each comparison did not suggest obvious publication bias but only two to five points were available for each analysis.

## Discussion

Specific criteria for the diagnosis of PCOS remain elusive. The clinical manifestations of hyperandrogenism may not always accompany elevated androgen levels since genetic susceptibility to these hormones at the target organ is also required. Hyperinsulinaemia has been documented to occur in 50–100% of obese PCOS patients and as high as 22% of lean PCOS patients (Dale *et al.*, 1992). Extreme cases of hyperinsulinaemia may present with hyperandrogenism, insulin resistance, and acanthosis nigricans (HAIR-AN syndrome). The most recent NIH consensus conference on PCOS in 1990 defined the syndrome as chronic anovulation with hyperandrogenism. LH:FSH ratios, ultrasound criteria and hyperinsulinaemia were hitherto not included in this definition. The recent 2003 Rotterdam ESHRE/ASRM consensus on PCOS attempted to delineate a clearer definition of PCOS

but also stated that 'PCOS remains a syndrome and no single diagnostic criterion (such as hyperandrogenism or PCO) is sufficient for clinical diagnosis. PCOS also remains a diagnosis of exclusion.' Nevertheless, they stated that the revised diagnostic criteria include two of the following three findings: (i) oligo- or anovulation, (ii) clinical and/or biochemical signs of hyperandrogenism, (iii) polycystic ovaries and exclusion of other aetiologies (ESHRE and ASRM, 2004). Clearly, as the Rotterdam proceedings suggested, the definition of PCOS may be flexible but the criteria employed in a study should directly address the interventions and outcomes assessed. The revised Rotterdam criteria are appropriate here. Ovulation and pregnancy rates, not metabolic parameters, were assessed in this study.

The evidence that hyperinsulinaemia causes hyperandrogenism, which in turn affects fertility, is more convincing than arguments that hyperandrogenism causes hyperinsulinaemia. Following insulin infusion, adrenal androgen levels rise (Elkind-Hirsch *et al.*, 1991). Conversely, treatment of hyperinsulinaemia with insulin-sensitizing agents or weight loss [5% or body mass index (BMI) <27 kg/m<sup>2</sup>] decreases androgen levels (Loverro *et al.*, 2002; Mitkov *et al.*, 2002). Medically or surgically induced menopause decreases circulating androgen levels without any effect on serum insulin levels (Nagamani *et al.*, 1986).

Previously, primary therapy PCOS patients involved ovarian drilling, possibly followed by ovulation induction with CC or gonadotrophins (Farquhar *et al.*, 2001). Ovarian drilling was hypothesized to decrease ovarian androgen production. Currently, standard treatment for ovulation induction in these patients is CC. The success rate of CC varies; 80% ovulation and 30–40% pregnancy rates have been reported (Gorlitsky *et al.*, 1978; Lunenfeld *et al.*, 1991; Kousta *et al.*, 1997; Imani *et al.*, 1998, 1999). Seventy-five per cent of pregnancies are achieved within the first 3 months of an ovulatory dose of CC (Imani *et al.*, 1998). Some authors have suggested that ovulation induction with CC for a period >6 months in properly selected patients with PCOS can provide a cumulative pregnancy rate >90% (Messinis and Milingos, 1997). The proposed mechanism of action is that CC is an anti-estrogen which leads to increased production of pituitary production of gonadotrophins (FSH and LH). CC does not address the hyperandrogenic or hyperinsulinaemic environment. CC also has an anti-estrogenic effect on the endometrial lining and cervical mucus. The incidence of multiple pregnancy (mostly twins) with CC is 4–10% (Kousta *et al.*, 1997; Eijkemans *et al.*, 2003). Although the 1% risk of ovarian hyperstimulation syndrome (OHSS) after gonadotrophin therapy may be much less with CC, it may still occur.

Metformin is a biguanide insulin sensitizer. It is labelled as class B in pregnancy. Sufficient human data are not available but the drug has not been associated with congenital defects in animals. It has been used in the treatment of women with diabetes mellitus II without negative effects on the fetus. Recently, small series of women who have taken metformin throughout pregnancy to prevent miscarriage and gestational diabetes have been published and there were no



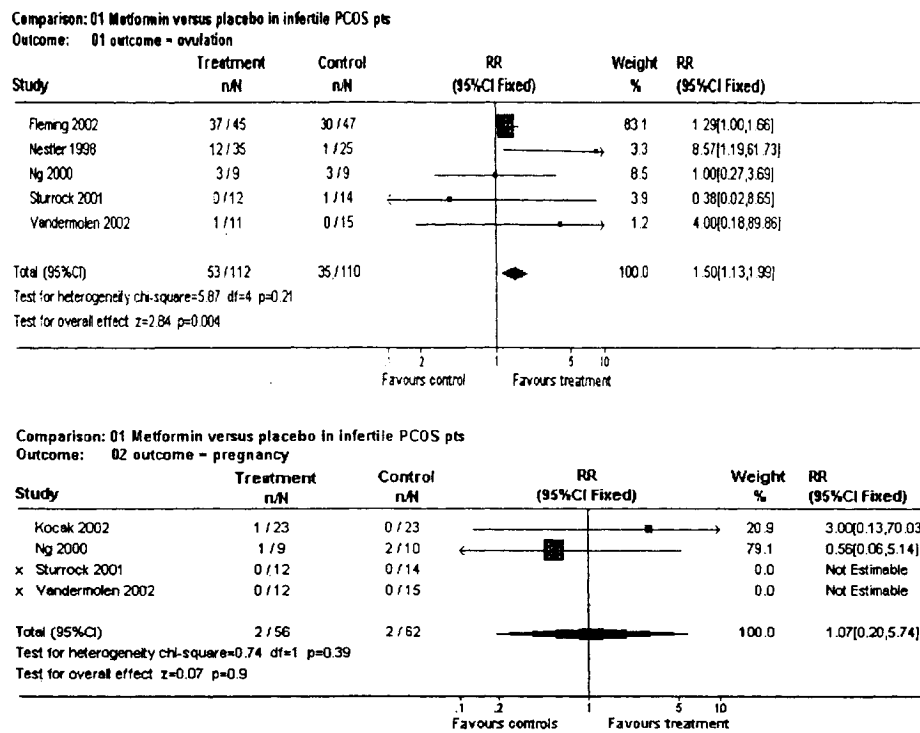


Figure 2. Comparison of metformin versus placebo in infertile polycystic ovarian syndrome (PCOS) patients. The upper panel shows the outcome ovulation and the lower panel the outcome pregnancy. RR = relative risk; CI = confidence interval. Generated from Meta-view 4.0.

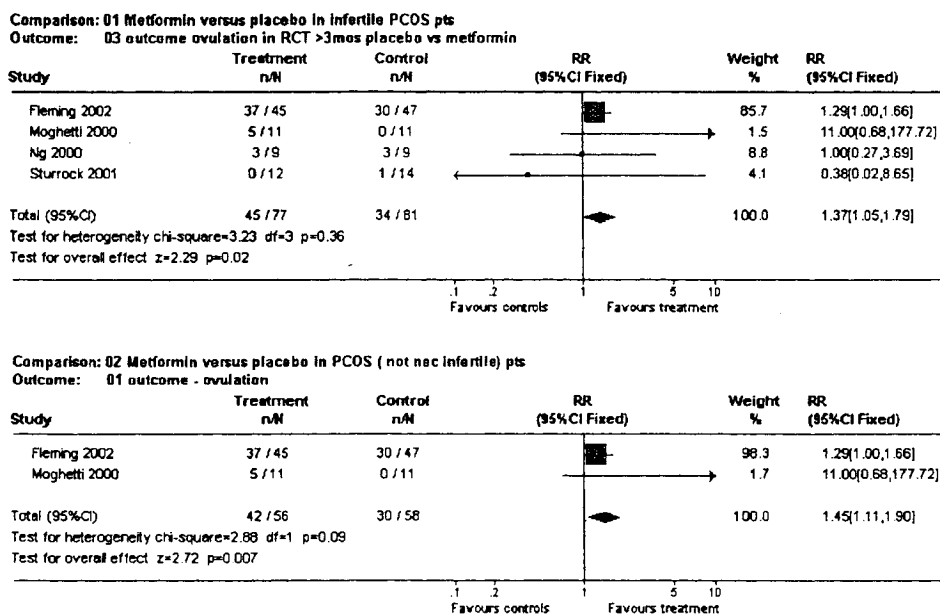


Figure 3. Comparison of metformin versus placebo with outcome ovulation in randomized treatment controlled trials > 3 months treatment duration. The upper panel shows infertile polycystic ovarian syndrome (PCOS) patients and the lower panel not necessarily (nec) infertile patients with PCOS. RR = relative risk; CI = confidence interval. Generated from Meta-view 4.0.

reported defects (Glueck *et al.*, 2002; Jakubowicz *et al.*, 2002). The mechanism of action is mainly the inhibition of hepatic gluconeogenesis. Metformin also increases peripheral glucose utilization and insulin sensitivity, but it is not associated with hypoglycaemia.

Recently, Lord *et al.* (2003) published meta-analyses with similar results in both the Cochrane Collaboration and the

*British Medical Journal*. The protocol for their paper was identified during our search but the papers were published after the last search date. While their results concerning ovulation and pregnancy were similar to those reported here, six studies were included in both papers, our paper included two different studies and their paper included five different studies (Nestler and Jakubowicz, 1996; Moggetti *et al.*, 2000;

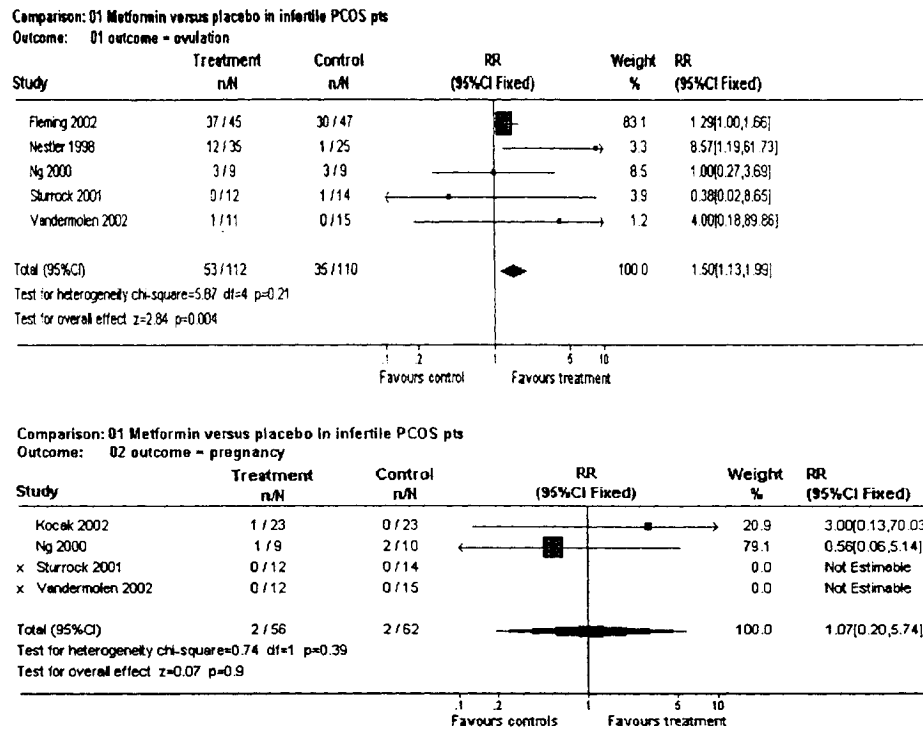


Figure 2. Comparison of metformin versus placebo in infertile polycystic ovarian syndrome (PCOS) patients. The upper panel shows the outcome ovulation and the lower panel the outcome pregnancy. RR = relative risk; CI = confidence interval. Generated from Meta-view 4.0.

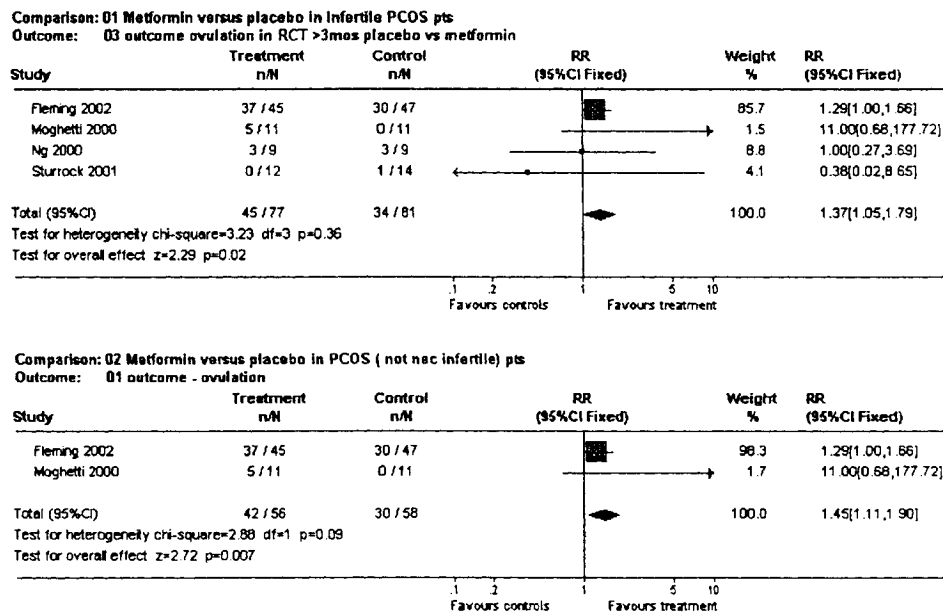


Figure 3. Comparison of metformin versus placebo with outcome ovulation in randomized treatment controlled trials > 3 months treatment duration. The upper panel shows infertile polycystic ovarian syndrome (PCOS) patients and the lower panel not necessarily (nec) infertile patients with PCOS. RR = relative risk; CI = confidence interval. Generated from Meta-view 4.0.

reported defects (Glueck *et al.*, 2002; Jakubowicz *et al.*, 2002). The mechanism of action is mainly the inhibition of hepatic gluconeogenesis. Metformin also increases peripheral glucose utilization and insulin sensitivity, but it is not associated with hypoglycaemia.

Recently, Lord *et al.* (2003) published meta-analyses with similar results in both the Cochrane Collaboration and the

British Medical Journal. The protocol for their paper was identified during our search but the papers were published after the last search date. While their results concerning ovulation and pregnancy were similar to those reported here, six studies were included in both papers, our paper included two different studies and their paper included five different studies (Nestler and Jakubowicz, 1996; Moggetti *et al.*, 2000;

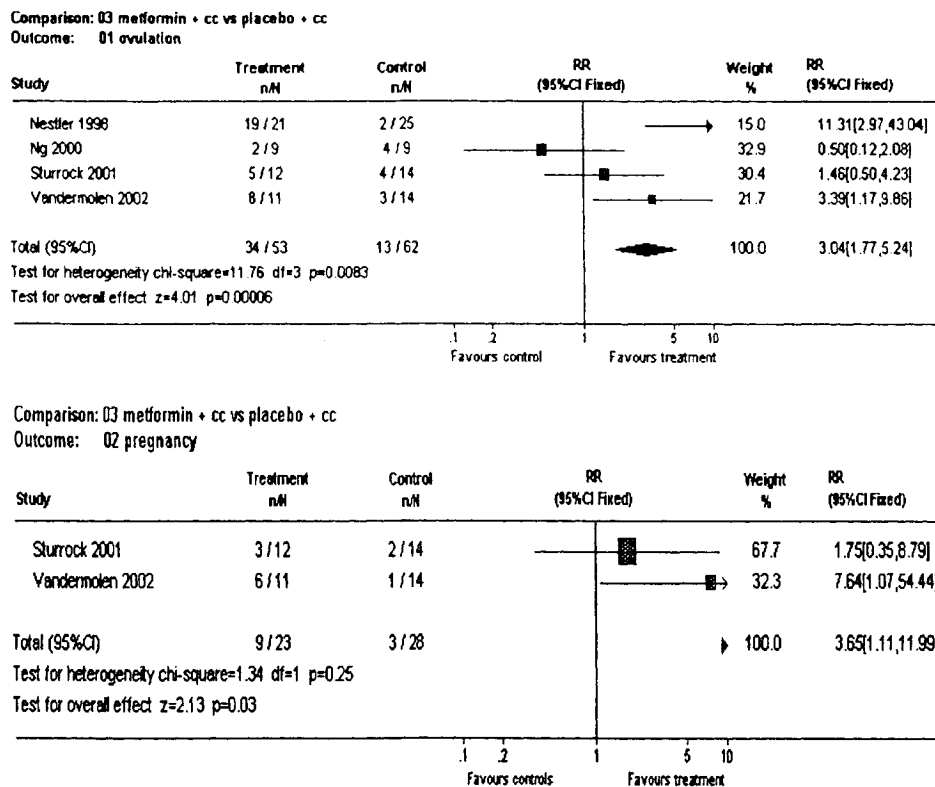


Figure 4. Comparison of metformin + clomiphene citrate (CC) versus placebo + CC in infertile patients. The upper panel shows the outcome ovulation and the lower panel the outcome pregnancy. RR = relative risk. Generated from Meta-view 4.0.

El-Biely and Habba, 2001; Jakubowicz *et al.*, 2001; Sturrock *et al.*, 2001; Malkawi and Qulban, 2002; Yarali *et al.*, 2002). Two of these five papers were not available to us (El-Biely and Habba, 2001; Malkawi and Qulban, 2002). Another evaluated co-administration of metformin during recombinant (r) FSH treatment of patients with CC-resistant PCOS (Yarali *et al.*, 2002). Two other papers had been initially identified but nothing in their title, abstract or methods suggested relevance (Nestler and Jakubowicz, 1996; Jakubowicz *et al.*, 2001). Nevertheless, our results are quite similar (Table IV).

Meta-analyses are limited by biases introduced through individual studies as well as biases introduced through the

processes of systematic review and quantitative summary. One consistent problem is that none of the studies were powered to assess the most important outcome, i.e. pregnancy. The follow-up varied from 1 to 6 months. While oligomenorrhoea/ovulation was generally a criterion, studies varied in their assessment and reporting of hyperandrogenaemia and CC resistance and this may have produced selection bias. The 'cross-over design' used in several studies to assess metformin plus CC versus CC was inappropriate. RCT also differed significantly in their reporting of methods used to randomize and in analysis methods (intention to treat versus per protocol). Contamination and cross-over were not reported. Not all studies evaluated infertile PCOS patients and when they did they did not always assess duration of infertility or other contributing factors (misclassification bias). We attempted to reduce publication, multiple publication, reporting, conformity and retrieval biases by comprehensive literature searches and searches for grey literature. Only one reviewer completed all steps so that extractor and recording bias was possible. We attempted to limit bias by extracting information on three separate occasions. RCT data are less prone to bias than observational data and by nature of the study design, differences in control versus study groups should balance once the study population has been defined.

No RCT directly compared metformin to CC for induction of ovulation and/or attainment of pregnancy. Cohort studies were insufficiently homogeneous to allow quantifiable summary. This review provides level 1a evidence regarding

Table IV. Metformin in polycystic ovarian syndrome (PCOS): comparison of meta-analyses results

	Kashyap <i>et al.</i> (in press) RR (95% CI)	Lord <i>et al.</i> (2003) OR (95% CI)
Ovulation		
Metformin versus placebo		
Infertile	1.50 (1.13, 1.99)	3.88 (2.25, 6.69)
Non-infertile	1.45 (1.11, 1.90)	
Metformin + CC	3.04 (1.77, 5.24)	4.41 (2.37, 8.22)
Pregnancy		
Metformin versus placebo*	1.07 (0.20, 5.74)	2.76 (0.85, 8.98)
Metformin + CC versus CC	3.65 (1.11, 11.99)	4.40 (1.96, 9.85)

\*Not significant.

RR = relative risk; OR odds ratio; CI = confidence interval; CC = clomiphene citrate; non-infertile = patients with PCOS who were not complaining of infertility.

metformin versus placebo for ovulation induction and pregnancy. This study suggests that metformin is superior to placebo for ovulation induction in patients with PCOS but that this benefit is not more pronounced with longer therapy (i.e. >3 months). The definition of PCOS is very important in determining the group of patients who will have an optimal response. Ng *et al.* (2001) did not find a benefit with metformin. The patients in their study were not overweight (BMI <23 kg/m<sup>2</sup>) and also were not hyperandrogenaemic as opposed to the other studies where a benefit was found. In a well-defined group of PCOS patients who do not complain of infertility, metformin also has significant advantage over placebo for resumption of ovulation and regulation of menstrual cycles.

Nevertheless, the data so far do not demonstrate a benefit of metformin versus placebo when the outcome considered is pregnancy. The follow-up time to pregnancy was short and, in the quantitatively summarized studies, pregnancy was not the primary outcome nor were these studies powered to assess pregnancy as an outcome. When we compared metformin plus CC to CC or metformin alone, however, there appeared to be a significant benefit of the combination treatment for both ovulation and pregnancy in patients with PCOS who were both hyperandrogenemic and overweight. However, the comparison of metformin with CC versus placebo with CC was done in each of these studies as a 'sequential study' rather than a true RCT or cross-over study.

A need still exists to compare directly metformin and CC as first-line agents for ovulation induction and achievement of singleton pregnancy in patients with well-defined PCOS. PCOS patients may gain more than ovulation induction from metformin therapy, including reduced miscarriage rates, lower incidences of multiple pregnancies, gestational diabetes, and ovarian hyperstimulation, and longer term cardiovascular health benefits. The currently ongoing study (Pregnancy in Polycystic Ovarian Syndrome: PPCOS) is designed to answer the first question and has the potential to provide valuable insight into the clinical management of this elusive syndrome.

## References

- Adams J, Polson JW, Franks S (1986) Prevalence of polycystic ovaries in women with anovulation and idiopathic hirsutism. *Br Med J* 293,355–359.
- Berlin JA (1997) Does blinding of readers affect the results of meta-analysis? *Lancet* 350,185–186.
- Clark HD, Wells GA, Huet C, McAlister FA, Salmi LR, Fergusson D and Laupacis A (1999) Assessing the reliability of randomized trials: reliability of the Jadad scale. *Controlled Clin Trials* 20,448–452.
- Dale P, Tanbor T and Vaaler S (1992) Body weight, hypersulinemia, and gonadotropin levels in the polycystic ovarian syndrome: evidence of two distinct populations. *Fertil Steril* 58,487–491.
- Eijkemans MJC, Imani B, Mulders AGMGJ, Habbema JDF and Fauser BCJM (2003) High singleton live birth rate following classical ovulation induction in normogonadotrophic anovulatory infertility (WHO 2). *Hum Reprod* 18,2357–2362.
- El-Biely MM and Habba M (2001) The use of metformin to augment the induction of ovulation in obese infertile patients with polycystic ovary syndrome. *Mid East Fertil Soc J* 6,43–49.
- Elkind-Hirsch K, Valdes C and McConnell T (1991) Androgen responses to acutely increased endogenous insulin levels in hyperandrogenic and normal cycling women. *Fertil Steril* 55,486.
- ESHRE and ASRM (2004) Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome\*1. *Fertil Steril* 81,19–25.
- Farquhar, C., Vandekerckhove, P., Lilford, R (2001) Laparoscopic "drilling" by diathermy or laser for ovulation induction in anovulatory polycystic ovary syndrome [update of Cochrane Database Syst Rev. 2000;(2): CD001122; PMID: 10796746]. *Cochrane Database of Systematic Reviews* CD001122.
- Fleming R, Hopkinson ZE, Wallace M, Greer IA and Sattar N (2002) Ovarian function and metabolic factors in women with oligomenorrhea treated with metformin in a randomized double blind placebo-controlled trial. *J Clin Endocrinol Metab* 87,569–574.
- Flight, N.R (2002) Insulin-sensitizing drugs for polycystic ovary syndrome (protocol). *The Cochrane Library*.
- Glueck C, Wang P, Fontaine R, Tracy T and Sieve-Smith L (1999) Metformin-induced resumption of normal menses in 39 of 43 (91%) previously amenorrheic women with polycystic ovarian syndrome. *Metab Clin Exp* 48,511–519.
- Glueck CJ, Wang P, Kobayashi S, Phillips H and Sieve-Smith L (2002) Metformin therapy throughout pregnancy reduces the development of gestational diabetes in women with polycystic ovary syndrome. *Fertil Steril* 77,520–525.
- Gorlitsky G, Kase N and Speroff L (1978) Ovulation and pregnancy rates with clomiphene citrate. *Obstet Gynecol* 51,265–269.
- Imani B, Eijkemans MJC, te Velde ER, Habbema JDF and Fauser BCJM (1998) Predictors of patients remaining anovulatory during clomiphene citrate induction of ovulation in normogonadotrophic oligoamenorrheic infertility. *J Clin Endocrinol Metab* 83,2361–2365.
- Imani B, Eijkemans MJC, te Velde ER, Habbema JDF and Fauser BCJM (1999) Predictors of chances to conceive in ovulatory patients during clomiphene citrate induction of ovulation in normogonadotrophic oligoamenorrheic infertility. *J Clin Endocrinol Metab* 84,1617–1622.
- Jadad AR, Moore A, Carroll D, Jenkinson C, Reynolds JM, Gavaghan D and McQuay HJ (1996) Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clin Trials* 17,1–12.
- Jadad AR, Moher D and Klassen TP (1998) Guides for reading and interpreting systematic reviews: II. How did the authors find the studies and assess their quality? *Arch Pediatr Adolesc Med* 152,812–817.
- Jakubowicz DJ, Seppala M, Jakubowicz S, Rodriguez-Armas O, Rivas-Santiago A, Koistinen H, Koistinen R and Nestler JE (2001) Insulin reduction with metformin increases luteal phase serum glycodeilin and insulin-like growth factor-binding protein 1 concentrations and enhances uterine vascularity and blood flow in the polycystic ovary syndrome. *J Clin Endocrinol Metab* 86,1126–1133.
- Jakubowicz DJ, Iuorno MJ, Jakubowicz S, Roberts KA and Nestler JE (2002) Effects of metformin on early pregnancy loss in the polycystic ovary syndrome. *J Clin Endocrinol Metab* 87,524–529.
- Juni P, Holenstein F, Sterne J, Bartlett C and Egger M (2002) Direction and impact of language bias in meta-analyses of controlled trials: empirical study [comment]. *Int J Epidemiol* 31,115–123.
- Kocak M, Caliskan E, Simsir C and Haberal A (2002) Metformin improves ovulatory rates, cervical scores, and pregnancy rates in clomiphene citrate-resistant women with polycystic ovary syndrome. *Fertil Steril* 77, 101–106.
- Kousta E, White DM and Franks S (1997) Modern use of clomiphene citrate in induction of ovulation. *Hum Reprod Update* 3,359.
- Lord JM, Flight IHK and Norman RJ (2003) Metformin in polycystic ovary syndrome: systematic review and meta-analysis. *Br Med J* 327,951–950.
- Loverro G, Lorusso F, De PagolaNicolardi G, Mei L and Selvaggi L (2002) Clinical and endocrinological effects of 6 months of metformin treatment in young hyperinsulinemic patients affected by polycystic ovarian syndrome. *Gynecol Endocrinol* 16,217–224.
- Lunenfeld B, Pariente C, Dor J, Menashe Y, Seppala M, Mortman H and Insler V (1991) Modern aspects of ovulation induction. *Ann NY Acad Sci* 626,207–216.
- Malkawi HY and Qulban HS (2002) The effect of metformin plus clomiphene citrate on ovulation and pregnancy rates in clomiphene-resistant women with polycystic ovary syndrome. *Saudi Med J* 23,663–666.
- McAuley L, Pham B, Tugwell P and Moher D (2000) Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 356,1228–1231.
- Messinis IE and Milingos SD (1997) Current and future status of ovulation induction in polycystic ovary syndrome. *Hum Reprod Update* 3,235–253.

- Mitkov M, Simenov S and Pehlivanov B (2002) Effect of metformin treatment of polycystic ovary syndrome. *Endocrinologia* 7,25–29.
- Moggetti P, Castello R, Negri C, Tosi F, Perrone F, Caputo M, Zanolin E and Muggeo M (2000) Metformin effects on clinical features, endocrine and metabolic profiles, and insulin sensitivity in polycystic ovary syndrome: a randomized, double-blind, placebo-controlled 6-month trial, followed by open, long-term clinical evaluation. *J Clin Endocrinol Metab* 85,139–145.
- Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, Liberati A, Linde K and Penna A (1996) Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. [comment]. *Lancet* 347,363–366.
- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D and Stroup DF (1999) Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of reporting of meta-analyses [comment]. *Lancet* 354,1896–1900.
- Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR and Liberati A (2000) What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol* 53,964–972.
- Nagamani M, Dinh T and Kolver M (1986) Hyperinsulinemia in hyperthecosis of the ovaries. *Am J Obstet Gynecol* 154,384.
- Nestler JE and Jakubowicz DJ (1996) Decreases in ovarian cytochrome P450c17{alpha} activity and serum free testosterone after reduction of insulin secretion in polycystic ovary syndrome. *N Engl J Med* 335, 617–623.
- Nestler JE, Jakubowicz DJ, Evans WS and Pasquali R (1998) Effects of metformin on spontaneous and clomiphene induced ovulation in the polycystic ovary syndrome. *N Engl J Med* 338,1876–1880.
- Ng EHY, Wat NMS and Ho PC (2001) Effects of metformin on ovulation rate, hormonal and metabolic profiles in women with clomiphene-resistant polycystic ovaries: a randomized, double-blind placebo-controlled trial. *Hum Reprod* 16,1625–1631.
- Singh I, Bedaiwy MA, Hatwal A, Kumar A and Agarwal A (2001) Increased pregnancy rates with metformin and clomiphene citrate in non-obese patients with polycystic ovary syndrome: prospective randomized study. *Fertil Steril* 76,S94.
- Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA and Thacker SB (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *J Am Med Assoc* 283,2008–2012.
- Sturrock NDC, Lannon B and Fay TN (2001) Metformin does not enhance ovulation induction in clomiphene resistant polycystic ovary syndrome in clinical practice. *Br J Pharmacol* 53,469–473.
- Vandermolen DT, Ratts VS, Evans WS, Stovall DW, Kauma SW and Nestler JE (2001) Metformin increases the ovulatory rate and pregnancy rate from clomiphene citrate in patients with polycystic ovary syndrome who are resistant to clomiphene citrate alone. *Fertil Steril* 75,310–315.
- Velazquez E, Mendoza S, Hamer T, Sosa F and Glueck C (1994) Metformin therapy in polycystic ovarian syndrome reduces hyperinsulinemia, insulin resistance, hyperandrogenemia, and systolic blood pressure, while facilitating normal menses and pregnancy. *Metab Clin Exp* 43,647–654.
- Yarali H, Yildiz BO, Demiroglu A, Zeyneloglu HB, Yigit N, Bukulmez O and Koray Z (2002) Co-administration of metformin during rFSH treatment in patients with clomiphene citrate-resistant polycystic ovarian syndrome: a prospective randomized trial. *Hum Reprod* 17,289–294.

Submitted on February 9, 2004; resubmitted on May 4, 2004; accepted on July 6, 2004

Appendix C2

Oxford Centre for Evidence-based Medicine Levels of Evidence (May 2001)

Level	Therapy/Prevention, Aetiology/Harm	Prognosis	Diagnosis	Differential diagnosis/symptom prevalence study	Economic and decision analyses
1a	SR (with homogeneity*) of RCTs	SR (with homogeneity*) of inception cohort studies; CDR† validated in different populations	SR (with homogeneity*) of Level 1 diagnostic studies; CDR† with 1b studies from different clinical centres	SR (with homogeneity*) of prospective cohort studies	SR (with homogeneity*) of Level 1 economic studies
1b	Individual RCT (with narrow Confidence Interval‡)	Individual inception cohort study with ≥ 80% follow-up; CDR† validated in a single population	Validating** cohort study with good††† reference standards; or CDR† tested within one clinical centre	Prospective cohort study with good follow-up****	Analysis based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multi-way sensitivity analyses
1c	All or none§	All or none case-series	Absolute SpPins and SnNouts††	All or none case-series	Absolute better-value or worse-value analyses ††††
2a	SR (with homogeneity*) of cohort studies	SR (with homogeneity*) of either retrospective cohort studies or untreated control groups in RCTs	SR (with homogeneity*) of Level >2 diagnostic studies	SR (with homogeneity*) of 2b and better studies	SR (with homogeneity*) of Level >2 economic studies
2b	Individual cohort study (including low quality RCT; e.g., <80% follow-up)	Retrospective cohort study or follow-up of untreated control patients in an RCT; Derivation of CDR† or validated on split-sample§§§ only	Exploratory** cohort study with good††† reference standards; CDR† after derivation, or validated only on split-sample§§§ or databases	Retrospective cohort study, or poor follow-up	Analysis based on clinically sensible costs or alternatives; limited review(s) of the evidence, or single studies; and including multi-way sensitivity analyses
2c	"Outcomes" Research; Ecological studies	"Outcomes" Research		Ecological studies	Audit or outcomes research
3a	SR (with homogeneity*) of case-control studies		SR (with homogeneity*) of 3b and better studies	SR (with homogeneity*) of 3b and better studies	SR (with homogeneity*) of 3b and better studies
3b	Individual Case-Control Study		Non-consecutive study; or without consistently applied reference standards	Non-consecutive cohort study, or very limited population	Analysis based on limited alternatives or costs, poor quality estimates of data, but including sensitivity analyses incorporating clinically sensible variations.
4	Case-series (and poor quality cohort and case-control studies§§)	Case-series (and poor quality prognostic cohort studies****)	Case-control study, poor or non-independent reference standard	Case-series or superseded reference standards	Analysis with no sensitivity analysis
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on economic theory or "first principles"

## Notes

Users can add a minus-sign "-" to denote the level of that fails to provide a conclusive answer because of:

- EITHER a single result with a wide Confidence Interval (such that, for example, an ARR in an RCT is not statistically significant but whose confidence intervals fail to exclude clinically important benefit or harm)
- OR a Systematic Review with troublesome (and statistically significant) heterogeneity.
- Such evidence is inconclusive, and therefore can only generate Grade D recommendations.

*	By homogeneity we mean a systematic review that is free of worrisome variations (heterogeneity) in the directions and degrees of results between individual studies. Not all systematic reviews with statistically significant heterogeneity need be worrisome, and not all worrisome heterogeneity need be statistically significant. As noted above, studies displaying worrisome heterogeneity should be tagged with a "*" at the end of their designated level.
†	Clinical Decision Rule. (These are algorithms or scoring systems which lead to a prognostic estimation or a diagnostic category.)
‡	See note #2 for advice on how to understand, rate and use trials or other studies with wide confidence intervals.
§	Met when all patients died before the Rx became available, but some now survive on it; or when some patients died before the Rx became available, but none now die on it.
\$\$	By poor quality cohort study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both exposed and non-exposed individuals and/or failed to identify or appropriately control known confounders and/or failed to carry out a sufficiently long and complete follow-up of patients. By poor quality case-control study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both cases and controls and/or failed to identify or appropriately control known confounders.
\$\$\$	Split-sample validation is achieved by collecting all the information in a single tranche, then artificially dividing this into "derivation" and "validation" samples.
††	An "Absolute SpPin" is a diagnostic finding whose Specificity is so high that a Positive result rules-in the diagnosis. An "Absolute SnNout" is a diagnostic finding whose Sensitivity is so high that a Negative result rules-out the diagnosis.
‡‡	Good, better, bad and worse refer to the comparisons between treatments in terms of their clinical risks and benefits.
†††	Good reference standards are independent of the test, and applied blindly or objectively to all patients. Poor reference standards are haphazardly applied, but still independent of the test. Use of a non-independent reference standard (where the 'test' is included in the 'reference', or where the 'testing' affects the 'reference') implies a level 4 study.
††††	Better-value treatments are clearly as good but cheaper, or better at the same or reduced cost. Worse-value treatments are as good and more expensive, or worse and the equally or more expensive.
**	Validating studies test the quality of a specific diagnostic test, based on prior evidence. An exploratory study collects information and trawls the data (e.g. using a regression analysis) to find which factors are 'significant'.
***	By poor quality prognostic cohort study we mean one in which sampling was biased in favour of patients who already had the target outcome, or the measurement of outcomes was accomplished in <80% of study patients, or outcomes were determined in an unblinded, non-objective way, or there was no correction for confounding factors.
****	Good follow-up in a differential diagnosis study is >80%, with adequate time for alternative diagnoses to emerge (eg 1-6 months acute, 1 - 5 years chronic)

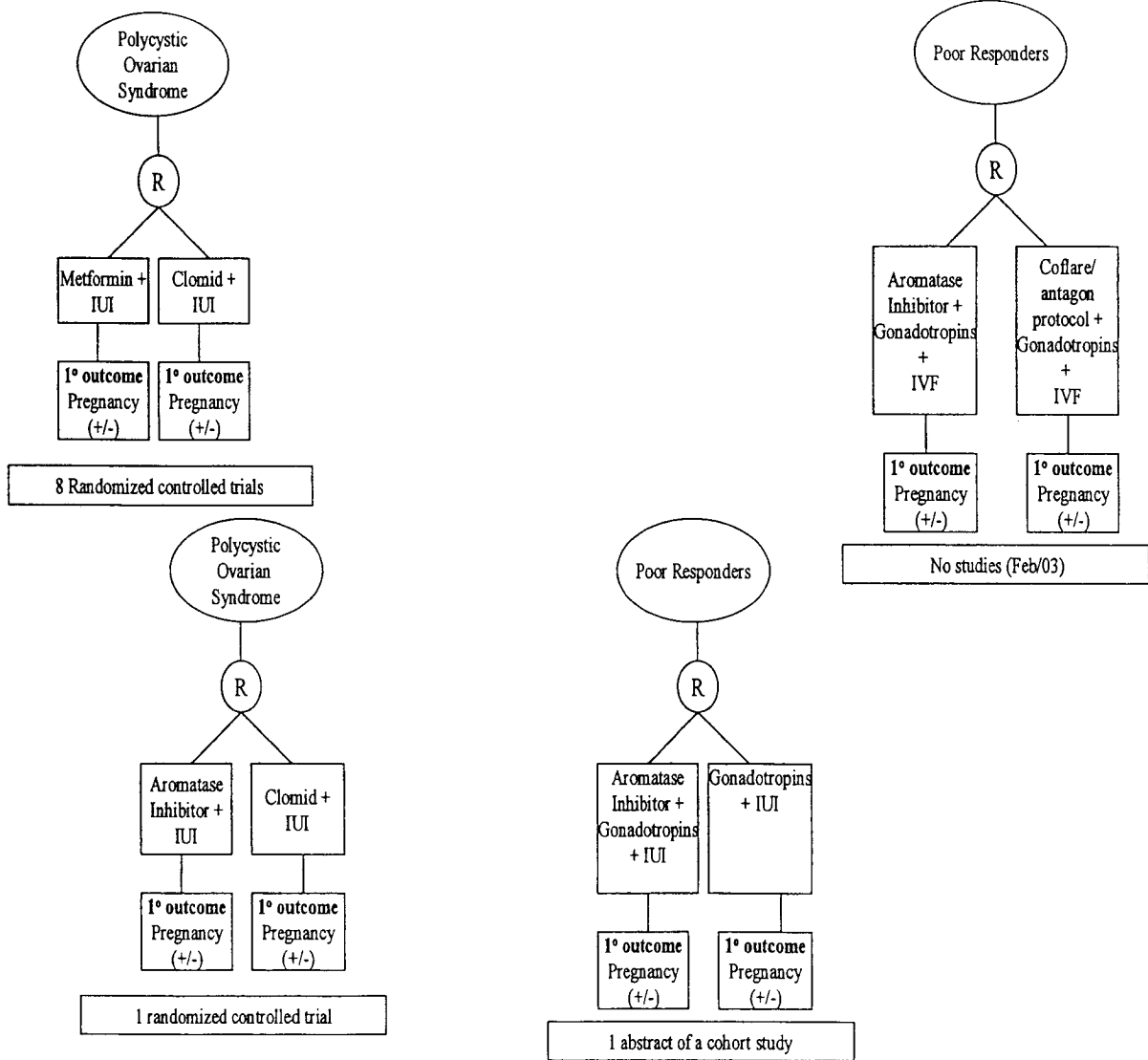
## Grades of Recommendation

A	consistent level 1 studies
B	consistent level 2 or 3 studies or extrapolations from level 1 studies
C	level 4 studies or extrapolations from level 2 or 3 studies
D	level 5 evidence or troublingly inconsistent or inconclusive studies of any level

"Extrapolations" are where data is used in a situation which has potentially clinically important differences than the original study situation.

## Appendix D

### Diagram illustrating change of RCT topic





**Appendix E**  
**Poor Responder Patient Consent**



Date (Mo., Dy., Yr)	Location	Service
Age	Doctor	

If No Plate, Print Name, Sex, And History No.

THE NEW YORK PRESBYTERIAN HOSPITAL-WEILL MEDICAL COLLEGE OF CORNELL  
UNIVERSITY

Consent Form for Clinical Investigation

**Project Title:** Aromatase inhibitors in IVF for Poor Responders  
**Subject:** \_\_\_\_\_ **Research Project #** \_\_\_\_\_

1. You are invited to participate in a research study of aromatase inhibitors for in-vitro fertilization. Physicians of The New York Presbyterian Hospital-Weill Medical College of Cornell University hope to learn if aromatase inhibitors – medications that can prevent early conversion of male hormone (androgens) to female hormones (estrogens) - are useful for patients who may otherwise have a poor response to stimulation medications. You were selected as a possible participant in this study because you have scheduled an in-vitro fertilization cycle at the Cornell Institute for Reproductive Medicine and have one or more of the following criteria Inclusion criteria:

- a) history of previous stimulation for IVF resulting in less than 5 follicles retrieved,
- b) estradiol pre-retrieval of < 1000 pg.ml in a previous cycle
- c) history of cancellation in previous IVF cycles for poor response.
- d) Age greater than 40 years old
- e) Previous high day 3 follicle stimulating hormone (FSH) or estradiol (E2)

The investigational drug, letrozole, is FDA (Food and drug administration) approved for use in women with advanced breast cancer. It decreases estrogen levels in these patients. Recently, several investigators have used it for ovarian stimulation in patients who will have artificial insemination or in-vitro fertilization. The addition of letrozole to ovarian stimulation

THE NEW YORK PRESBYTERIAN HOSPITAL (continued) (Use other side if necessary)

medications (gonadotropins) is the only way that the study patients differ from the standard treatments used in the study control group.

If you decide to participate and are eligible to participate (by inclusion/exclusion criteria), we will randomize (like the flip of a coin) you to either a protocol that includes letrozole (an aromatase inhibitor) and gonadotropin (stimulation) medication or to an alternative, standard protocol decided by your primary physician. The maximum duration of time you will be observed as part of the study is 10 weeks from the time of signing the consent to observation of pregnancy by ultrasound. This amount of time and the treatments undertaken does not differ in any way from how you would be clinically managed if you were not a participant in this study except that patients in the study group will use a medication called letrozole in addition to their stimulation medications. Different physicians have a different preferences for either of the treatments but not one of these treatments have been shown to be better than the other. This study will hope to answer if these treatments are equally acceptable or if one is better than the other. Patients who receive letrozole will take it orally every day for 5 days from day 2 –7 of their menses. They will then overlap gonadotropin medications, as prescribed by their primary physician, on day 5. Patients who receive the standard treatment will proceed as instructed by their physician. All patients will be monitored for follicular (egg) development with ultrasound and blood tests (approximately 1 tablespoon each time to test for estrogen levels) as would be normally done if they were not participants in this study. The frequency of blood tests and ultrasound is approximately every 1 –2 days during stimulation (total of approximately 12 blood tests and ultrasounds) and then as clinically indicated.

2. The costs for treatment will be billed to you as we routinely do in clinical practice. Subjects do not receive any compensation for their participation. Participation requires the following:
- a. a special visit with the study team to ensure that you meet the inclusion and exclusion criteria and to sign this consent.
  - b. At this visit you will be randomized to the letrozole treatment or the standard treatment. We will give you the prescription for the letrozole. The cost of letrozole may vary by pharmacy but is approximately \$10 per pill. You will require 5 pills (total ~\$50 – 100.). The cost of the letrozole arm will not be significantly more or less expensive than standard treatment. The cost of medications is the responsibility of the patient.
  - c. You will then come in for routine blood test (approximately 1 tablespoon) on day 2 of your menses. IF these tests are normal (FSH and estradiol) then you will receive a phone call that night to start your medication. Both groups will be treated the same from here on in.
  - d. As per standard at our center, monitoring, with transvaginal ultrasounds and blood tests (approximately 1 tablespoon), is done every 1 –2 days during stimulation. Once you begin stimulation, you will follow the routine procedures. Human Chorionic gonadotropin (hCG) will be given once follicles are approximately 17-18 mm, as is standard. Retrieval will follow approximately 34 hours later. The method of fertilization for eggs will be determined clinically by your history ( ie

intracytoplasmic sperm injection –ICSI versus insemination). Patients will then return for embryo transfer on day 3 post-retrieval.

3. Your participation in the project involves the following risks:
  - a) the risks from participating in this study are minimal and not different from the risks encountered from these standard treatments associated with IVF if you were not participating in this study.
  - b) The risks to the patients are those of the risks of IVF in general: the risk of infection or bleeding secondary to the oocyte retrieval (0.3 – 3 %); the risk of injury to abdominal organs or blood vessels during the procedure (less than 1%); and the risk of anesthetic for the retrieval procedure (less than 0.1%). These risks are not increased as a result of the study since all of the subjects have already planned to do IVF. The risk of ovarian hyperstimulation and or multiple pregnancy (Twins or triplets) is extremely low (less than 2%) in this specific patient group whose pregnancy rate is estimated to be lower than 10%. Again, this is the risk associated with IVF and should not be increased by the study. The use of letrozole has been shown to be associated with nausea and vomiting (11%). Longer term effects such as decreased bone density and adverse effects in the lipid profile have only been seen with long term treatment (greater than 3 months and 60 days respectively). Our patients would only use the medication for a total of 5 days.
  - c) Other uncommon side-effects from prolonged use of Letrozole include : headaches (13%), fatigue (11%), muscular aches, hot flushes (6%), and hair thinning. All of these effects are uncommon and are associated with longer use of letrozole than in this study. Also, all side effects return to normal with discontinuation of medication.
  
4. We cannot and do not guarantee that you will receive any benefits from this study. Pregnancy is the goal of treatment as it would be with or without the study. The general benefits to society are that the question as to whether or not letrozole is useful for ovarian stimulation in poor responders will hopefully be answered.
  - a) The treatments described are the standard, first line therapies used for patients with your histories with the exception of the addition of letrozole to the study group. Your clinical management will not be different than if you were not part of this study.
  - b) Reasons that we might discontinue from the study include the following: if you develop a concurrent illness, particularly liver or kidney illnesses, that may interfere with drug metabolism. If you develop side effects of the treatment that are persistent, we may have to discontinue you from the study.
  
5. Any information obtained during this study and identified with you will remain confidential and will be disclosed only with your permission. The exception to this is that the Institutional