Journal of Obstetrics and Gynaecology, October 2011; 31: 566–571

ISSN 0144-3615 print/ISSN 1364-6893 online
DOI: 10.3109/01443615.2011.596956

informa
healthcare

REVIEW

Is ultrasound monitoring of the ovaries during ovulation induction by clomiphene citrate essential? A systematic review

N. Galazis[1], M. Zertalis[1], Z. Haoula[2] & W. Atiomo[2]

[1]Nottingham Medical School and [2]Division of Human Development, School of Clinical Sciences, University of Nottingham, Queens Medical Centre Campus Nottingham University Hospitals, Nottingham, UK

The study objective was to investigate whether ultrasound (US) monitoring is essential during treatment with clomiphene citrate (CC) for ovulation induction, as recommended by the Royal College of Obstetricians and Gynaecologists (RCOG) and the National Institute for Clinical Excellence (NICE). We performed a systematic review of all studies investigating the effects of US in the treatment of ovulatory dysfunction with CC. The main objective of this review was to investigate whether US monitoring during CC treatment reduced multiple pregnancy rates. There was insufficient evidence to suggest that US monitoring reduces multiple pregnancy rates or improves pregnancy rates. On the other hand, no indication that treatment with CC is safe without US monitoring was identified. The small number of relevant studies and the heterogeneity observed in the methodologies of each study prohibit reliable conclusions to be drawn. There is currently no basis for amending the evidence base (good-practice points) used in the RCOG and NICE guidelines, which recommend the use of US to monitor the ovaries during stimulation with CC.

Keywords: Clomiphene citrate, ovulation induction, ovulatory dysfunction, systematic review, ultrasound

Introduction

Clomiphene citrate (CC), a selective oestrogen receptor modulator, is the mainstay of ovulation induction in women with ovulatory dysfunction or anovulation (e.g. due to polycystic ovary syndrome, PCOS). It achieves this by increasing the production of gonadotrophins, leading to ovarian stimulation and sometimes hyperstimulation, hence inducing ovulation. Unfortunately, ovarian hyperstimulation is associated with increased risks of multiple pregnancies, which are in turn associated with higher perinatal mortality and morbidity (Dickey 2007; Levene et al. 1992; Scialli 1986).

NICE and RCOG have recommended that patients treated with CC should be offered ultrasound (US) monitoring during at least the first menstrual cycle, to ensure they receive the right dose for inducing ovulation and thus minimise the risk of multiple pregnancies (RCOG 1998; NICE 2004). However, due to the high costs involved with US monitoring, not all women treated with CC are monitored. Furthermore, the evidence that universal monitoring is cost-effective is debatable, as the RCOG guideline on US monitoring in women on CC is based on good practice points (view of guideline development group) and not grade A (systematic reviews and meta-analysis of randomised controlled trials).

The aim of this systematic review was to determine whether US monitoring in women with ovulatory dysfunction or anovulation treated with CC reduced multiple pregnancy rates. This was done in order to objectively inform the current clinical guidelines. Another objective of this review was to determine if US monitoring of CC treatment improved pregnancy rates by improving the timing of intercourse. We further aimed to identify areas for future research.

Materials and methods

Studies eligible for review

The selection criteria were non-restrictive and included non-randomised-controlled trials (non-RCTs) because of the low number of relevant studies. Primary studies (written in English) comparing US monitoring against other monitoring techniques or against no monitoring at all were eligible for screening. The studies included for this review were finally selected by carefully screening their abstracts and methodology.

Finding relevant studies

The MEDLINE (1966–December 2010); EMBASE (1980–December 2010); CINAHL (1977–December 2010) and Cochrane (1993–December 2010) databases were searched using the terms 'anovulation'; 'ovulatory dysfunction'; 'ovulation induction'; 'clomiphene citrate'; clomifene citrate' and 'ultrasound monitoring', without any limits/restrictions. A manual search of references from all the studies was also conducted to identify any other potentially relevant studies. The search criterion ended in February 2011. The search findings were independently double-checked by one of the co-authors (ZH).

Methodological quality assessment

The methodological quality of the studies was determined by their internal validity, which depended on the following three factors: randomisation, presence of relevant outcome measures and description of an acceptable withdrawal/drop-out rate (up to 15% of the participants). The evaluation of other important quality parameters such as allocation concealment and blinding of the assessors was not feasible because of the nature of the experiments involving an invasive transvaginal ultrasound (TVUS). Furthermore, the selected studies were assessed for their descriptive and statistical validities (Figure 1).

Main outcome measures

The primary outcome measure was multiple pregnancy rates in CC treatment cycles monitored by US compared with no, or other forms, of monitoring. Secondary outcome measures included pregnancy rates (PR), the number of follicles stimulated by CC and ovarian hyperstimulation syndrome (OHSS) rates.

Correspondence: N. Galazis, Nottingham Medical School, University of Nottingham, Queens Medical Centre Campus Nottingham University Hospitals, Nottingham NG7 2UH, UK. E-mail: mzybng@nottingham.ac.uk

A study was classified as high quality (HQ) if it scored 3/3 for the internal validity. Studies scoring 2/3 or less were classified as low quality (LQ). Descriptive validity of each study was determined by checking if the groups were similar at baseline and that description of any adverse effects, duration of the trial, index and control interventions and selection criteria was adequate. Furthermore, statistical validity was assessed by checking that the sample size was described and that point estimates and measures of variability were included in the results. As with searching for the relevant studies, methodological quality assessment was also carried out independently by another reviewer (MZ).

Figure 1. Methodological quality assessment.

Results

Selecting the relevant studies

Figure 2 demonstrates the selection process of the relevant papers. The initial literature search, which was conducted through MEDLINE, yielded 59 relevant papers. From these, 48 were excluded by reading their title and abstract. These papers were irrelevant to our study objective as they used, e.g. other ovulation-inducing pharmaceuticals for clomiphene-resistant cases. No review was found that specifically explored the effects of US as a monitoring modality in women treated with CC. Two papers were excluded by carefully reading their research aims and methodology. In these cases, US and CC were used, however, the aim of these studies was to evaluate the effectiveness of CC in different patient groups, rather than investigating the effects of US monitoring in these groups. One study (Kousta et al. 1997) emphasised the importance of US monitoring during treatment with CC in order to choose the appropriate dose of CC in subsequent cycles and minimise the risks of multiple pregnancy and OHSS. However, no relevant data were presented on US monitoring during CC treatment in the paper. For that reason, this study was excluded from this systematic review. Finally, three papers were excluded from the review because their full texts were not written in English. Searching through the Cochrane, CINAHL and EMBASE databases and hand searching of the references of relevant manuscripts did not yield any additional papers. This left five primary studies eligible for this review.

General characteristics of studies

Five papers identified from the literature, involving 380 participants, were considered relevant to this review (Abdul-Karim et al. 1990; Coughlan et al. 2010; Giannopoulos et al. 2005; Smith et al. 1998; Zreik et al. 1999). Three of them compared US monitoring for ovulation induction against different monitoring modalities (Smith et al. 1998; Zreik et al. 1999) or against no monitoring (Abdul-Karim et al. 1990). Two of these studies were RCTs. Randomisation methods were adequately described in these two RCTs. These involved the use of an opaque envelope with assignments determined by a computer-generated random number table (Smith et al. 1998; Zreik et al. 1999). Abdul-Karim et al. (1990) performed a retrospective analysis. The remaining two papers attempted to compare the effects of US as a way of monitoring ovulation during CC treatment in different groups of infertile/subfertile women (e.g. with PCOS, unexplained infertility, ovulation dysfunction, etc.) to determine whether US monitoring could be selective to a specific pathological group instead of everyone with conception difficulties. These were either retrospective (Coughlan et al. 2010) or prospective studies (Giannopoulos et al. 2005). The number of participants (NOP) ranged between 40 and 182 (mean 76) in the different studies.

While all five studies described the inclusion criteria for the participants, common to all being anovulation (mainly due to PCOS), only one study presented data on their exclusion criteria (Smith et al. 1998). In general, the selection criteria varied between each study. Furthermore, the number of ovulatory cycles used in each study fluctuated from 1–2 (Giannopoulos et al. 2005), to a study with a median of six cycles (Abdul-Karim et al. 1990). One out of the five studies did not describe the dosage of CC administered to their patients (Abdul-Karim et al. 1990). From the four remaining study groups who described the dosage of CC used, two used 50–100 mg (Giannopoulos et al. 2005; Zreik et al. 1999); one used 25–50 mg (Coughlan et al. 2010) and one used 50–150 mg (Smith et al. 1998).

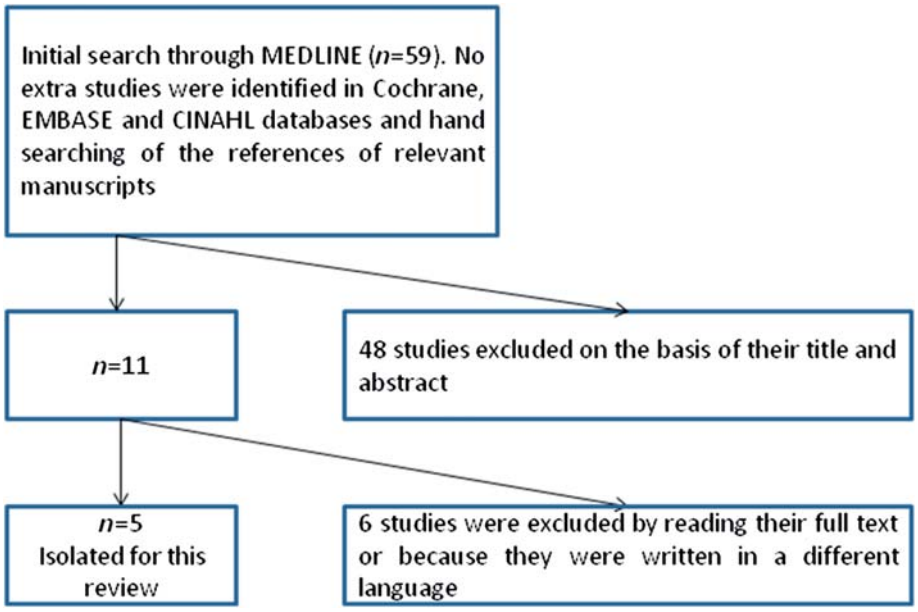

Figure 2. A flow chart summarising the selection process.

Table I summarises the main characteristics and Table II presents the main outcome measures and findings of the aforementioned papers.

Table III illustrates the methodological quality of each study. Three out of five studies scored <3/3 for internal validity and were hence classified as low quality (LQ) (Abdul-Karim et al. 1990; Coughlan et al. 2010; Giannopoulos et al. 2005). The remaining two studies scored 3/3 and were classified as high quality (HQ) (Smith et al. 1998; Zreik et al. 1999). All but one explicitly described the index and control interventions used in the trials as well as the selection criteria of their participants (Coughlan et al. 2010). All studies described the duration of the trials. With regards to statistical validity, all five studies described their sample sizes but only two of them presented their results using point estimates and measures of variability (e.g. *p* values, standard deviations, etc.; Abdul-Karim et al. 1990 and Zreik et al. 1999).

The main outcome measures

There were insufficient data (Table II) to draw any conclusions on the primary outcome measure of this systematic review (multiple pregnancy rates in CC treatment cycles monitored by US compared with no, or other forms of monitoring). PR per cycle was the most consistent outcome measure used in the reviewed papers. In the two RCTs, there were no significant differences in the PR of women monitored with US compared with other methods of monitoring ovulation during CC treatment (Smith et al. 1998; Zreik et al. 1999).

In Coughlan et al. (2010), patients received 50 mg of CC in 52 of the 58 cycles and 25 mg in the remaining six. They reported a

Table I. The main characteristics of the studies included in the review.

Study	Study design	NOP	Age range (years)	Duration	Inclusion criteria	Exclusion criteria	NOC	Dose of CC	Ways of monitoring
Smith et al. (1998)	RCT	45	24–42	3 cycles	Anovulation, Oligo-ovulation, Luteal phase deficiency. Normal hystero-salpingogram, Normal serum [prolactin], Normal serum [TSH], Partner had normal semen analysis	Previous use of CC	3	50 mg on days 5–9. If no ovulation induction, 100 mg on days 5–9 of next cycle. If no ovulation, 150 mg on days 5–9 of next cycle	Group I: BBT only Group II: TVUS, BBT and urinary LH surge
Coughlan et al. (2010)	RS	182	n/a	2 years	Ovulatory dysfunction (including unexplained infertility) Regular ovulatory cycles receiving super-ovulation treatment as part of an IUI treatment	Not mentioned	2.3*	25 mg; 50 mg	TVUS No non-US group. Included because of US use
Giannopoulos et al. (2005)	PS	40	21–41	n/a	Anovulatory (including PCOS), Unexplained infertility	Not mentioned	1–2	50 mg on days 2–6. If no ovulation, 100 mg on days 2–6 of next cycle	TVUS, Luteal phase progesterone assays
Zreik et al. (1999)	RCT	54	33*	22 months	Unexplained infertility, Anovulation, Male factor infertility. Normal hysterosalpingogram, normal endometrial biopsy, and a history of CC use of <6 months' duration	Not mentioned	4	CC on days 3–7. Unexplained infertility: 100 mg/day Male factor infertility: 50 mg/day Anovulatory: initially received 50 mg/day increased by 50 mg per cycle if no ovulation was observed	Group A: urinary LH Group B: TVUS and administration of hCG when follicle was >18 mm
Abdul-Karim et al. (1990)	RS	59	NM	NM	Anovulation due to PCOS	n/a	Median 6	CC or hMG. Those who failed to conceive with CC were started on hMG. Dose n/a	US against no US

*Mean. NOP, number of participants; NOC, number of cycles; RCT, randomised-controlled trial; RS, retrospective study; PS, prospective study; BBT, basal body temperature; NM, not mentioned.

Table II. The main outcome measures and characteristics of the included studies.*

Study	Main outcome measures	Conclusions
Smith et al. (1998)	Pregnancy rates per cycle: Group I: 10%; Group II: 6% Multiple births: No data Number of follicles: No data OHSS: No data	No statistical significance between pregnancy rates in the two groups
Coughlan et al. (2010)	Pregnancy rates per cycle: 10.8% Number of >14 mm follicles in a cycle: 0, 9%; 1, 49%; 2, 28%; 3, 11%; 4, 3%† Multiple births: Twin, 4.3%; Triplet, 2.2% of all pregnancies OHSS: No data	Women receiving CC should be monitored with US to reduce multiple pregnancy rates
Giannopoulos et al. (2005)	Pregnancy rates: No data Multiple births: No data Number of follicles: 1, 62.5%; 2, 5%; 3, 2.5%‡ OHSS: 0%	Selective monitoring should be offered to women with unexplained infertility receiving CC as universal monitoring is rarely feasible due to lack of resources and financial cost
Zreik et al. (1999)	Pregnancy rate per cycle: Unexplained infertility; Group A, 9.5%; Group B, 2.3% Anovulation: Group A and B, 0% Male factor infertility: Group A, 0%; Group B, 11.1% Birth rates: No data Number of follicles: No data OHSS: No data	Timing IUI with the use of expensive and time-consuming method such as TVUS monitoring of folliculogenes and hCG induction of ovulation does not appear to produce an increased pregnancy rate over urinary LH monitoring of ovulation
Abdul-Karim et al. (1990)	Pregnancy rates: 5.9% per cycle§ Birth rates: 74% of pregnancies or 4% of participants Multiple births: No data OHSS: No data	US does not exert an adverse influence on the conception rate or conceptus viability when performed prior to the ovulation inducing stimulus. Overall, the advantages of POUS in determining the optimal time for ovulation induction outweigh its possible disadvantages.

OHSS, ovarian hyperstimulation syndrome; IUI, intrauterine insemination; POUS, peri-ovulatory ultrasound.
*Main outcome measure = pregnancy rates except for Giannopoulos et al. (2005).
†Incidence of ≥3 follicles over 14 mm was higher in normo-ovulatory than the ovulatory dysfunction group (17% vs 6%).
‡All patients stimulating three follicles were in the unexplained fertility group.
§Pregnancy rates varied with age and whether US was used or not: < 30 years: 9.8% vs > 30 years: 1.7%, US: 8.5%, no US: 4.5%.

PR per cycle of 10.8%. A total of 9% of the cycles treated with CC did not develop a follicle; 49% developed one follicle and 28% developed two follicles; 11% of women developed three follicles and 3% developed four follicles, despite receiving lower doses of CC than the standard recommended dose. Some 67% of the pregnancies resulted from cycles where one follicle was detected on US. Two of these mono-follicular cycles resulted in twins and one resulted in a triplet pregnancy. A total of 28% of the pregnancies in the Coughlan et al. (2010) study, resulted from cycles where two follicles were identified. These pregnancies were all singleton. The remaining 5% of the pregnancies in the same study occurred in cycles where three follicles were detected, and were all singleton.

Table III. The assessment of methodological quality of the selected studies.

		Internal validity				Descriptive validity					Statistical validity	
n	Study	A	B	C	Score	D	E	F	G	H	I	J
1	Smith et al. (1998)	Y	Y	Y	3/3, HQ	N	N	Y	Y	Y	Y	N
2	Coughlan et al. (2010)	N	Y	N	1/3, LQ	N	N	Y	N	N	Y	N
3	Giannopoulos et al. (2005)*	N	Y	N	1/3, LQ	N	N	Y	Y	Y	Y	N
4	Zreik et al. (1999)	Y	Y	Y	3/3, HQ	N	N	Y	Y	Y	Y	Y
5	Abdul-Karim et al. (1990)	N	Y	N	1/3, LQ	Y	N	Y	Y	Y	Y	Y

A, adequate randomisation; B, relevant outcome measures; C, withdrawal/drop-out rate described and acceptable; D, groups similar at baseline. E, adverse effects described; F, duration of trial described; G, index and control interventions explicitly described; H, selection criteria described; I, sample size in each group described; J, point estimates and measures of variability and presented for primary outcome measures; N, No or not applicable. Y, Yes. HQ, high quality; LQ, low quality.
*Included for the use of US monitoring rather than for the outcome measure.

Abdul-Karim et al. (1990) compared the PR of women receiving CC or human menopausal gonadotrophin (hMG) with or without US monitoring of ovulation. The PR per cycle when US was used was 8.5% compared with 4.5% when no US was used. However, these results are inconclusive due to the inconsistency in the treatment used (whether CC or hMG). Of these pregnancies, 74% resulted in births, however no data were found on multiple pregnancy rates.

Giannopoulos et al. (2005) tried to establish whether monitoring during ovulation induction with CC by means of TVUS and luteal phase progesterone assays could be reserved in specific groups of infertile women. TVUS identified one follicle in 62.5% of the cycles; two follicles in 5% and three follicles (of which all were in the unexplained infertility group), in 2.5% of the cycles. There was no ovulation in 30% of the cycles. Of all the studies in this review, this was the only group that presented data on OHSS which was 0%. There were no data on PR, birth rates and multiple pregnancies. With the exception of Giannopoulos et al. (2005) and Coughlan et al. (2010), the studies included for this review did not provide data on the number of follicles detected on each cycle treated with CC.

Discussion

This was the first systematic review of the literature attempting to evaluate the need for US monitoring during treatment of infertility/subfertility with CC. There were insufficient data to draw any

conclusions on the primary outcome measure of this systematic review, which was to investigate multiple pregnancy rates in CC treatment cycles monitored using US compared to no, or other forms, of monitoring. PR per cycle was the most consistent outcome measure used in the reviewed papers. In the two RCTs included, there were no significant differences in the PR of women monitored with US compared with other methods of monitoring ovulation during CC treatment (Smith et al. 1998; Zreik et al. 1999).

Searching through the major medical databases, namely MEDLINE, EMBASE, CINAHL and Cochrane, as well as hand-searching of the references of the relevant manuscripts, yielded only five studies suitable for this review. Only three of these directly compared US for ovulation induction against different modalities, with only two of these being RCTs (the gold-standard method for scientific research) (Smith et al. 1998; Zreik et al. 1999). More direct comparison of US against other monitoring methods (e.g. BBT, luteal progesterone assays, etc.) in the form of RCTs are imperative before consistent conclusions can be drawn regarding the necessity of US during treatment of infertility with CC.

Heterogeneity among the studies was identified. This was mainly due to differences in population studied and outcome measures. To improve accuracy, future studies need to aim for a set of objectives, standardised designs and outcome measures. Furthermore, the dose of CC should always be reported and standardised. This was not always the case in the existing studies. In addition, larger study sample groups should be aimed for in future studies. This should be based on proper power calculations, based on a specific outcome measure, e.g. multiple pregnancy rates.

When US was directly compared against a different monitoring technique or against no technique, no evidence was found to suggest that multiple pregnancy rates were reduced or that OHSS was prevented when US monitoring took place. In fact, data on OHSS (0%) were reported in only one study (Giannopoulos et al. 2005). OHSS can potentially be a severe complication of fertility medication and it therefore needs to be reported in future research. Evidence that US monitoring prevents this could be decisive in supporting its use during CC treatment.

There was, however, some evidence supporting the selective use of US monitoring in certain subgroups of infertile women to minimise multiple pregnancy rates. Such potential groups included women with 'unexplained infertility' that have been shown to produce a greater number of follicles during CC treatment (Giannopoulos et al. 2005). However, these results must be reproduced in future studies in order for reliable conclusions to be drawn. This was particularly highlighted by the findings in this systematic review, which found that there was not always a direct correlation between the number of follicles identified on a US scan and the multiple birth rates.

Coughlan et al. (2010) highlighted these possible inaccuracies of US monitoring. Women identified as having only one follicle during US follicle tracking ended-up with multiple pregnancies and singleton pregnancies occurred in cycles where two follicles were identified. On the other hand, as some cycles where multi-follicular development was noted were cancelled in this study (Coughlan et al. 2010), this systematic review has not provided sufficient data to recommend that it is safe to monitor CC without US at this point in time. The issues above also highlight the fact that meticulous training and auditing needs to be carried out by the healthcare personnel who perform and report on US scans during CC treatment cycles.

The main strength of this systematic review was that it is the first one focusing on the effects of US for monitoring the ovaries during ovulation induction with CC. To minimise any potential bias, the selection of the relevant papers as well as methodological quality assessment were carried out independently by two authors. In this way, the most relevant studies were selected from the literature. The qualitative analysis performed in this review should enable the reader to make an informed judgement on the necessity of US during treatment of women with CC. The methodological assessment also allowed us to evaluate the validity of the information obtained from each study.

On the other hand, this review had a few weaknesses, mainly arising because the small number of studies currently available did not allow us to perform a meta-analysis. Only two out of the five studies were RCTs which, according to the criteria set in this review, were of high methodological quality (HQ). That, along with the heterogeneity observed in the primary studies, was an important limitation to our objective of determining the necessity of US for monitoring ovulation induction.

In conclusion, this review found insufficient data to draw any conclusions to support the NICE and RCOG guidelines which recommend that CC treatment cycles in infertile women should be monitored using US to reduce multiple pregnancy rates. We however, cannot recommend that US monitoring during treatment with CC is stopped based on this review, as there was also insufficient evidence to suggest that this would be safe. In other words, there is currently no basis on which to amend the strength of the current evidence base (good-practice points) used in the RCOG and NICE guidelines on the need for US monitoring during CC treatment in infertile women. More studies are however required to address this issue, given the current cost implications of routine US monitoring and the invasiveness of the practice to fertility patients. We do however, envisage potential ethical challenges with conducting a direct comparison of US monitoring to no monitoring or a non-imaging based monitoring approach (e.g. luteal phase progesterone assays) in infertile women undergoing CC treatment because of the risks of missing higher order births in the non-monitoring group. There is, however, a need for more research to identify the subgroup of infertile women where US monitoring may not be extensively required, in order to minimise costs and make the process of conception more natural (and less invasive) for couples.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Abdul-Kari[n]m RW, Terry FM, Badawy SZ, Sheehe PR. 1990. Effect of ultrasound monitoring of follicular growth on the conception rate. A clinical study. Journal of Reproductive Medicine 35:147–151.

Coughlan C, Fitzgerald J, Milne P, Wingfield M. 2010. Is it safe to prescribe clomiphene citrate without ultrasound monitoring facilities? Journal of Obstetrics and Gynaecology 30:393–396.

Dickey RP. 2007. The relative contribution of assisted reproductive technologies and ovulation induction to multiple births in the United States 5 years after the Society for Assisted Reproductive Technology/American Society for Reproductive Medicine recommendation to limit the number of embryos transferred. Fertility and Sterility 88:1554–1561.

Giannopoulos T, Sherriff E, Croucher C. 2005. Follicle tracking of women receiving clomiphene citrate for ovulation induction. Journal of Obstetrics and Gynaecology 25:169–171.

Kousta E, White DM, Franks S. 1997. Modern use of clomiphene citrate in induction of ovulation. Human Reproduction Update 3:359–365.

Levene MI, Wild J, Steer P. 1992. Higher multiple births and the modern management of infertility in Britain. The British Association of Perinatal Medicine. British Journal of Obstetrics and Gynaecology 99:607–613.

NICE. 2004. Clinical Guideline No. 11. Fertility: assessment and treatment for people with fertility problems. London: National Institute for Clinical Excellence.

RCOG. 1998. National evidence-based clinical guidelines. The management of fertility in secondary care. London: Royal College of Obstetricians and Gynaecologists.

Smith YR, Randolph JF Jr, Christman GM, Ansbacher R, Howe DM, Hurd WW. 1998. Comparison of low-technology and high-technology monitoring of clomiphene citrate ovulation induction. Fertility and Sterility 70:165–168.

Scialli AR. 1986. The reproductive toxicity of ovulation induction. Fertility and Sterility 45:315–323.

Zreik TG, García-Velasco JA, Habboosh MS, Olive DL, Arici A. 1999. Prospective, randomized, crossover study to evaluate the benefit of human chorionic gonadotropin-timed versus urinary luteinizing hormone-timed intrauterine inseminations in clomiphene citrate-stimulated treatment cycles. Fertility and Sterility 71:1070–1074.