What is the most valid comparison treatment in trials of intrauterine insemination, timed or uninfluenced intercourse? A systematic review and meta-analysis of indirect evidence

H.K. Snick¹,5, J.A. Collins²,3,6 and J.L.H. Evers⁴,7

¹Department of Obstetrics and Gynaecology, Ziekenhuis Walcheren, Vlissingen, The Netherlands; ²Faculty of Health Sciences, McMaster University, Hamilton, Canada; ³Faculty of Medicine, Dalhousie University, Halifax, Canada; ⁴Department of Obstetrics and Gynaecology, Centre for Reproductive Medicine and Biology, GROW Research Institute, Maastricht University Medical Centre, Faculty of Health, Medicine and Life Sciences, Maastricht University, PO Box 5800, 6202AZ Maastricht, The Netherlands; ⁵Present address: 78 bd. Tixador 66140, Canet en Roussillon, France; ⁶Present address: 400 Mader’s Cove Road, RR 1, Mahone Bay, Canada NS B0J 2E0
⁷Correspondence address. E-mail: jev@sgyn.azm.nl

BACKGROUND: Timed intercourse (TI), which is the usual control treatment in trials of intrauterine insemination (IUI), is not a typical coital activity and could impair fertility. This review summarizes the trials of IUI of male partner’s prepared semen among subfertile couples according to whether the control group had TI or expectant management. METHODS: A search of relevant databases and bibliographies until February 2008 yielded 150 citations of which 31 were potentially relevant and 11 met all criteria. The total estimates of the differences in pregnancy rates per couple were calculated with weights equal to the inverse variance. The primary analysis was a categorical meta-analysis by the type of control treatment (TI or expectant management). RESULTS: In 11 trials with 13 comparisons of IUI and intercourse among 1329 couples with subfertility, the average difference in pregnancy rate between IUI and controls was 6.1% in trials with TI and 3.9% in trials with expectant management, as the control. The adjusted indirect estimate of the difference between the types of control groups was 2.8% (95% CI 2.6, 10.7). The difference by type of control treatment was not significant, neither in the 11 most relevant trials (P = 0.82), nor in a broader group of 19 trials and 2512 patients (P = 0.20). CONCLUSIONS: The additional benefit accruing to IUI, where TI is the control, is not significant, but it is consistent with the possibility that pregnancy may be less likely in TI controls than in expectant management controls.

Keywords: intrauterine insemination; timed intercourse; subfertility; ovarian stimulation; meta-analysis

Introduction

Intrauterine insemination (IUI) of prepared sperm is a common treatment for subfertility, which is often combined with ovarian stimulation (OS), using either clomiphene citrate or gonadotrophins (Verhulst et al., 2006; Bensdorp et al., 2007). In the Netherlands, there were ~28 500 cycles of IUI in 2003, and in France 42 843 cycles of IUI were reported for 2000 (Commenges-Ducos, 2004; Steures et al., 2007c). In six trials involving 517 women with unexplained subfertility, who were receiving OS, there was a moderate benefit from IUI (OR 1.62, 95% confidence interval (CI) 1.13, 2.50) compared with timed intercourse (TI) (Verhulst et al., 2006). In male subfertility, IUI was not significantly superior to TI in natural cycles OR 5.3, 95% CI 0.42–67) or in cycles with OS (OR 1.67, 95% CI 0.83–3.37) (Bensdorp et al., 2007).

In the majority of trials, the control treatment for IUI is TI. This choice appears to be founded on a belief that a credible comparison of IUI with intercourse requires that intercourse should be timed in a manner similar to the timing of IUI. Although this comparison is explanatory, showing that IUI can be superior to TI that is not a typical coital activity. Thus, the comparison is not pragmatic and could not show whether IUI does work better in typical practice. Notably, several recent multicentre studies of IUI have not adopted TI as the control treatment, preferring expectant management, which allows couples to make their own choices with respect to the timing and frequency of intercourse (Bhattacharya et al., 2006; Steures et al., 2006; Steures et al., 2007a).

Because TI involves interfering with natural coital habits by refraining from intercourse until some marker shows that ovulation is imminent, it may reduce the likelihood of pregnancy
Selection of studies
Relevance was determined from titles and then from abstracts. Full reports were reviewed for all potentially relevant citations. Where more than one study were reported from the same centre, only the most recent study was accepted. Studies were further evaluated to determine whether the co-interventions used in the IUI and intercourse arms were equivalent; whether the allocation was truly randomized and whether in crossover study designs, outcome data were available for the first phase of the trial.

Data extraction
Data were abstracted from eligible studies on country, allocation, study design, treatment type, co-interventions (clomiphene citrate, gonadotrophin), details of IUI procedures, number of patients and cycles and number of pregnancies or births. If the data were not presented in the form of a 2 × 2 table, a table was re-constructed from the available data.

Analysis
Weighted average summary estimates of rate differences in pregnancy rates per couple were calculated with weights equal to the inverse variance. Between-study heterogeneity was estimated with the use of the Q statistic and was judged to be excessive when $P < 0.10$. Heterogeneity was further quantified by $I^2$, the proportion of variability across studies that is due to heterogeneity rather than chance (Higgins et al., 2003). Fixed and random effects models were not materially different; random effects models are presented because heterogeneity could not be ruled out in most analyses (Deeks et al., 2001).

The primary meta-analysis was a categorical analysis of the most relevant trials by type of control treatment (TI or expectant management). This categorical analysis partitioned the heterogeneity into the components from the model (control group type) and the residual heterogeneity, an analysis which is analogous to the partitioning of variance in ANOVA (Hedges and Olkin, 1985). Adjusted indirect comparisons were used to estimate differences between groups (Glenny et al., 2005); these differences were estimated for summary rate differences and for summary relative rates.

Other categorical analyses evaluated separately the effects of diagnostic group (unexplained, male factor, cervical factor, not categorized) and type of OS (none, clomiphene citrate, gonadotrophin). Two sensitivity analyses were planned which added excluded studies to the primary categorical analysis. The first of these added back the crossover studies that lacked results after the first phase; the second sensitivity analysis added studies that involved alternate allocation. A final exploratory sensitivity analysis removed outlier studies from the primary analysis.

Publication bias was explored by funnel plot and analysed by the Begg and Mazumdar rank correlation test (Begg and Mazumdar, 1994; Egger et al., 1997).

Results
Study selection
From the 150 PubMed citations and additional sources, 31 studies were reviewed in full (Fig. 1). Five studies had a different co-intervention for IUI and control cycles, which would confound the comparison of control groups: three involved TI (Deaton et al., 1990; Ho et al., 1992; Aribarg and Sukcharoen, 1995) and two involved expectant management (Tummon et al., 1997; Steures et al., 2006). No studies were found that used OS protocols and expectant management. Seven studies

Methods
Search strategy
The search strategy used the terms (IUI [Title/Abstract]) AND (pregnancy [Title/Abstract]) AND randomized [Title/abstract] (150 citations). The last search was in February 2008. Studies were eligible if they (i) compared IUI with intercourse, (ii) had pregnancy or live birth per couple among the reported outcomes and (iii) reported pregnancy results in a manner that allowed the creation of two-by-two tables from the study data. Meeting abstracts were accepted if they included all data necessary for our analysis. The search was not restricted by language and it included Medline and Embase citations together with additional studies that were identified from the reference lists of the retrieved articles. The literature was searched independently by two authors (H.S., J.C.), and no disagreement occurred. Minor discrepancies were solved by consensus involving all three authors. Furthermore, the Cochrane Menstrual Disorder and Subfertility Group Trials Register was searched, along with the Cochrane Central Register of Controlled Trials (CENTRAL) in the Cochrane Library, the ISRCTN Register, the abstracts of the ASRM and ESHRE annual meetings from 1998 to 2007, and the references of the six published meta-analyses on IUI (Hughes et al., 1997; Zeyneloglu et al., 1998; Bensdorp et al., 2007; Cohen et al., 1999; Helmerhorst et al., 2006; Verhulst et al., 2007).
Figure 1: Flow chart. Possibly relevant studies, reasons for exclusion from analysis and remaining eligible studies.

with TI as the comparator presented pregnancy data only by cycle and not by patient (Kerin et al., 1984; Kerin and Quinn, 1987; Ho et al., 1989; te Velde et al., 1989; Evans et al., 1991; Martinez et al., 1991; Zikopoulos et al., 1993). Three studies had inadequate randomization, two with TI as the comparator (Check and Spirito, 1995; Janko et al., 1998) and one with expectant management (Glazener et al., 1987). Pre-crossover data were not available for a further five studies (Doyle and DeCherney, 1991; Kirby et al., 1991; Crosignani and Walters, 1994; Gregoriou et al., 1995; Jaroudi et al., 1998).

Studies included

Eleven studies remained for the primary analysis (Table I). Three involved no co-intervention (Martinez et al., 1990; Bhattacharya et al., 2006; Steures et al., 2007b), two involved clomiphene citrate co-intervention in IUI and intercourse groups (Karlstrom et al., 1993; Agarwal and Mittal, 2004) and seven made use of gonadotrophin OS for IUI and intercourse (Crosignani et al., 1991; Karlstrom et al., 1993; Nan et al., 1995; Melis et al., 1995; Arcaini et al., 1996; Gregoriou et al., 1995). Only two studies, both without co-intervention, had expectant management as the control intercourse option (Bhattacharya et al., 2006; Steures et al., 2007b). None of the studies found a means of achieving blinding for intervention. The 11 studies involved 13 comparisons and 1279 patients: Karlstrom et al. (1993) stratified the treatment and control groups by co-intervention (clomiphene citrate, gonadotrophin treatment); Melis et al. (1995) stratified the groups by diagnosis (unexplained, male factor).

Preliminary meta-analyses

Pregnancy data and individual study rate differences are shown in Table II and with the summary estimate in Fig. 2. The pregnancy rate differences between IUI and intercourse ranged from −22.4 to 23.8%. The heterogeneity Q value was 26.7, with 12 degrees of freedom, \( P = 0.009. \) In 13 estimates from 11 studies, the random effects summary rate difference was 6.0% in favour of IUI (95% CI −0.9, 12.6).

One preliminary categorical meta-analysis indicated that the heterogeneity was not associated with diagnosis of unexplained subfertility or male factor (\( P = 0.46. \) In another preliminary categorical meta-analysis, the average rate differences for no stimulation, clomiphene citrate and gonadotrophin studies, respectively, were \( 7.4 (0.3, 26.3), -17.7 (-22.4, -11.5) \) and \( 9.8% (4.0, 16.6), (P = 0.012). \) Although type of co-intervention contributed significantly to the heterogeneity among studies, the trend did not reflect the relative strength of clomiphene citrate and gonadotrophins.

Primary categorical meta-analysis

The primary categorical meta-analysis was a random effects model involving 13 estimates from 11 studies. The average pregnancy rate difference in 11 estimates from nine studies comparing IUI and TI was 6.1% (95% CI −2.2, 14.6). The average of two estimates from two studies comparing IUI and expectant management was 3.9% (2.8, 4.7). The \( P \)-value for between groups heterogeneity was 0.82, indicating that the 2.2% (6.1 − 3.9) average difference between studies with TI and expectant management controls was not significant (Table III). Adjusted indirect comparisons. The adjusted indirect estimate of the difference between average pregnancy rates with TI and with expectant management controls was 2.2% (95% CI −6.3, 10.7). The relative likelihood of pregnancy with IUI was 1.25 (95% CI 0.87, 1.86) in studies with TI controls and 1.23 (95% CI 1.15, 1.28) in studies with expectant management controls. The adjusted indirect estimate of the apparent superiority of relative rates with TI controls compared with expectant management controls was 1.01% (95% CI 0.69, 1.49).

Sensitivity analyses

Sensitivity analyses were planned to test the judgments made about the eligibility of studies for analysis. The first sensitivity

<table>
<thead>
<tr>
<th>Authors</th>
<th>Contrast</th>
<th>Allocation concealment</th>
<th>Diagnosis</th>
<th>Co-intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agarwal and Mittal (2004)</td>
<td>parallel</td>
<td>ns (not stated)</td>
<td>Unexplained</td>
<td>CC</td>
<td>TI</td>
</tr>
<tr>
<td>Arcaini et al. (1996)</td>
<td>parallel</td>
<td>ns</td>
<td>Unexplained</td>
<td>G</td>
<td>TI</td>
</tr>
<tr>
<td>Bhattacharya et al. (2006)</td>
<td>parallel</td>
<td>Remote</td>
<td>Unexplained</td>
<td>None</td>
<td>EM</td>
</tr>
<tr>
<td>Chung et al. (1995)</td>
<td>parallel</td>
<td>Sealed envelopes</td>
<td>Unexplained</td>
<td>G</td>
<td>TI</td>
</tr>
<tr>
<td>Crosignani et al. (1991)</td>
<td>crossover</td>
<td>ns</td>
<td>Male</td>
<td>G</td>
<td>TI</td>
</tr>
<tr>
<td>Gregoriou et al. (1996)</td>
<td>crossover</td>
<td>ns</td>
<td>Male, mild endometriosis</td>
<td>CC, G</td>
<td>TI</td>
</tr>
<tr>
<td>Karlstrom et al. (1993)</td>
<td>parallel</td>
<td>ns</td>
<td>Unexplained, male</td>
<td>None</td>
<td>TI</td>
</tr>
<tr>
<td>Martinez et al. (1990)</td>
<td>Latin square</td>
<td>ns</td>
<td>Unexplained</td>
<td>Male</td>
<td>G</td>
</tr>
<tr>
<td>Melis et al. (1995)</td>
<td>parallel</td>
<td>sealed envelopes</td>
<td>Unexplained</td>
<td>G</td>
<td>TI</td>
</tr>
<tr>
<td>Nan et al. (1994)</td>
<td>crossover</td>
<td>sealed envelopes</td>
<td>Male</td>
<td>G</td>
<td>TI</td>
</tr>
<tr>
<td>Steures et al. (2007b)</td>
<td>parallel</td>
<td>sealed envelopes</td>
<td>Cervical</td>
<td>None</td>
<td>EM</td>
</tr>
</tbody>
</table>

CC, clomiphene citrate; G, gonadotrophins; TI, timed intercourse; EM, expectant management.
analysis added seven comparisons of IUI and intercourse involving a further 866 patients from five crossover studies that did not report pregnancy rate for the first phase before crossover, a total of 20 comparisons from 16 studies involving 2195 patients. In the random effects model, the average pregnancy rate difference for the 18 estimates comparing IUI and TI was 8.9% (95% CI 2.8, 14.1). The average of two estimates comparing IUI and expectant management was 4.0% (2.8, 4.7). The 4.9% (8.9 − 4.0) difference between studies with TI and expectant management controls was associated with P = 0.51.

The second sensitivity analysis added three more comparisons of IUI and intercourse involving a further 317 patients from three studies that used alternate allocation, for a total of 23 comparisons from 19 studies involving 2512 patients. In the random effects model, the average pregnancy rate difference for the 20 estimates comparing IUI and TI was 9.6% (95% CI 4.2, 14.4). The average of three estimates comparing IUI and expectant management was 1.9% (−2.1, 4.7). The 7.7% (9.6 − 1.9) difference between studies with TI and expectant management controls was associated with P = 0.20.

A final exploratory sensitivity analysis was conducted that modified the primary analysis by omitting the two estimates which involved clomiphene citrate co-intervention. Reference to Fig. 2 shows the clomiphene estimates were associated with the most negative results for IUI. In the analysis of 11 comparisons from 10 studies involving 1126 patients, heterogeneity was no longer significant (Q = 10.9, 10 degrees of freedom, P = 0.36). In a fixed effects model, the average
pregnancy rate difference for the nine estimates comparing IUI and TI was 11.3% (95% CI 6.0, 17.1). The average of two estimates comparing IUI and expectant management was 4.3% (2.8, 4.7). The adjusted indirect difference between studies with TI and expectant management controls was 7.0 (95% CI −4.6, 18.7) (P = 0.15).

**Publication bias**

For the 11 studies in the primary analysis, the funnel plot and the rank-correlation test (z = 0.061; P = 0.54) indicated that there was no significant publication bias. For the 16 studies in the first sensitivity analysis and the 19 studies in the second sensitivity analysis, the corresponding P values are 0.90 and 0.73.

**Discussion**

In this systematic review of trials comparing IUI and intercourse among couples with subfertility, the difference between IUI and control pregnancy rates in trials with TI as the control was 6.1% in favour of IUI, compared with 3.9% in trials with expectant management as the control. The difference between trials with the two types of control treatment was not significant, however, neither in the 11 most relevant trials (P = 0.82), nor in a broader group of 19 trials (P = 0.20). Although non-significance could imply lack of a true difference, the analyses also are consistent with the possibility that pregnancy is less likely using TI than with expectant management.

The strengths of the present study include the broad inclusion criteria (improving the generalizability of its findings) and the standard methods of assembly and analysis. The meta-analysis made use of rate differences rather than ORs or relative risks, because rates are directly relevant to clinical practice. The quality of any systematic review depends, however, on the methodology of the trials, and IUI trial methodology on average is sub-optimal, due to the lack of information on allocation concealment and randomization procedures (Verhulst et al., 2006). The primary meta-analysis necessarily included studies with different co-interventions because no study with OS of both intervention groups also had expectant management as the control. It is relevant to consider whether the difference between IUI and control of any kind would be greater with gonadotrophin co-intervention. To address this point, a relative risk meta-analysis corresponding to the primary categorical meta-analysis was done. For this comparison, the difference between control types was also not significant, and the magnitude of the difference is small whether in absolute (2.8%) or relative terms (relative likelihood 1.01).

Five studies had to be excluded because a different co-intervention for IUI and control cycles would have confounded the comparison of control groups so that these studies could not possibly be relevant to the question (Deaton et al., 1990; Ho et al., 1992; Ari barg and Sukcharoen, 1995; Tummon et al., 1997; Steures et al., 2006). A further seven studies were excluded because they lacked pregnancy rates per patient (Kerin et al., 1984; Kerin and Quinn, 1987; Ho et al., 1989; te Velde et al., 1989; Evans et al., 1991; Martinez et al., 1991; Zikopoulos et al., 1993). The seven studies might have been included had we been willing to combine pregnancy rates per cycle and per patient. We chose not to add this group of studies because they were uninformative with respect to UI and therefore did not justify a further complexity in the analysis.

In trials assessing IUI compared with controls, the added rate difference with TI rather than expectant management as the control intervention appears to be small, but if significant it would be clinically important. The potentially spurious TI effect accounts for 36% [(6.1 − 3.9)/6.1] of what has been assumed to be the benefit of IUI. In the exploratory analysis after excluding the clomiphene citrate treatment arms that caused heterogeneity, the TI effect accounted for 62% of the IUI benefit. Of course, a non-significant difference between two effect estimates means either that there is no true difference, in which case the apparent difference arose merely by chance, or that there is a true difference and the trial lacked the power needed to estimate the difference with appropriate confidence. This analysis of all relevant trials involved 485 and 794 patients in the expectant and TI groups, respectively. If it was an individual trial, there would be only 44% power to evaluate a difference between 6.1 and 3.9% pregnancy rates per person, assuming alpha, two-tailed, was 5%.

This review showed some pitfalls of comparing traditional, much-used clinical interventions like IUI and TI, especially when the appropriateness of the comparison group is under discussion. TI may be preferable to UI as control for IUI in efficacy trials, however, in ‘real life’ effectiveness trials UI is the most appropriate comparator. Both should be regarded with care however. In TI cycles, how compliant were the couples, did they refrain from intercourse until the moment of TI? In UI cycles, what was the frequency of intercourse (being as strongly related as it is to duration of subfertility and age of the couple), and what was the resulting time interval between intercourse and ovulation, or between intercourse and optimal mucus quality? Wilcox et al. (2004), when surveying women with non-hormonal contraception, found the frequency of spontaneous intercourse to increase during the follicular phase of the cycle, peaking at the time of ovulation, and declining abruptly thereafter. Does the same pattern exist in couples whose sex life has been disrupted for months or even years by a fertility investigation?

It is clear that in an effectiveness or pragmatic analysis, TI, if timed according to an external marker, e.g. the LH surge, is not an appropriate comparator for IUI. Under these circumstances TI may not be as effective as spontaneous uninfluenced coital activity (Snick, 2005). Wilcox et al. (1995) found the chances of conception to be highest on the 2 days before and the day of ovulation, rapidly decreasing to almost zero on the day after ovulation. The present study, providing additional suggestive evidence that pregnancy may be less likely with TI than with expectant management, is in agreement with this and is also in line with previous work by Nulsen et al. (1987) who found a rapid decline of cervical mucus quality in the 24 h following the LH surge. This rapidly increasing cervical hostility may perhaps not constitute an immediate
problem in case of normospermia, it may become of importance in couples with sperm defects or unexplained subfertility. In the latter, the fertile window may be shortened (Keulers et al., 2007), and it may close before ovulation, especially in ovulation induction patients, who are at risk of premature luteinization, reflected in poor mucus quality resulting in premature closure of their fertile mucus window. Premature poor mucus quality may be overcome by IUI, but it may turn into a barrier to conception in spontaneous or, especially, TI.

In conclusion, although the present review showed that the additional benefit accruing to IUI where TI is the control is not significant and could reflect lack of a true difference, it is also consistent with the possibility that pregnancy may be less likely in TI controls than expectant management controls. One reason for a decrease in conception rate with TI as compared with UI may be the decreasing cervical mucus quality after the LH surge. This may constitute a barrier to the spontaneous passage of sperm into the uterus and decrease conception chances if the couple has refrained from intercourse until after the LH surge has occurred. This is in contrast with expectantly managed couples who at that same moment will already have seized the opportunity and taken advantage of the favourable mucus when it was still present. TI is an artificial, and possibly even deleterious, alternative to UI.

Author roles
H.K.S. conceived the idea for the study, searched and abstracted the literature, and contributed to the manuscript and its revisions. J.A.C. searched and abstracted the literature, conducted the initial analyses and contributed to the manuscript and its revisions. J.L.H.E. advised on the methods, and contributed to the manuscript and its revisions.

References


Tummon IS, Asher LJ, Martin JSB, Tulandi T. Randomized controlled trial of superovulation and insemination for infertility associated with minimal or mild endometriosis. *Fertil Steril* 1997;68:8–12.


Submitted on May 5, 2008; resubmitted on May 5, 2008; accepted on May 9, 2008