Cycle cancellation and pregnancy after luteal estradiol priming in women defined as poor responders: a systematic review and meta-analysis

Kasey A. Reynolds1,*, Kenan R. Omurtag1, Patricia T. Jimenez2, Julie S. Rhee1, Method G. Tuuli1, and Emily S. Jungheim1

1Department of Obstetrics and Gynecology, Washington University School of Medicine, 4444 Forest Park Ave., Ste. 3100, St Louis, MO 63108, USA 2Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, TX, USA

*Correspondence address. Tel: +314-286-2400; Fax: +314-286-2455; E-mail: reynoldsk@wudosis.wustl.edu

Submitted on April 11, 2013; resubmitted on June 13, 2013; accepted on July 2, 2013

STUDY QUESTION: Does a luteal estradiol (LE) stimulation protocol improve outcomes in poor responders to IVF?

SUMMARY ANSWER: LE priming is associated with decreased cycle cancellation and increased chance of clinical pregnancy in poor responders

WHAT IS KNOWN ALREADY: Poor responders to IVF are one of the most challenging patient populations to treat. Many standard protocols currently exist for stimulating these patients but all have failed to improve outcomes.

STUDY DESIGN, SIZE, DURATION: Systematic review and meta-analysis including eight published studies comparing assisted reproduction technology (ART) outcomes in poor responders exposed to controlled ovarian hyperstimulation with and without LE priming. A search of the databases MEDLINE, EMBASE and PUBMED was carried out for studies in the English language published up to January 2012.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Studies evaluating women defined as poor responders to ART were evaluated. These studies were identified following a systematic review of the literature and data were analyzed using the DerSimonian–Laird random effects model. The main outcomes of interest were cycle cancellation rate and clinical pregnancy. Although the definition of clinical pregnancy varied between studies, the principal definition included fetal cardiac activity as assessed by transvaginal ultrasonography after 5 weeks of gestation.

MAIN RESULTS AND THE ROLE OF CHANCE: A total of 2249 publications were identified from the initial search, and the bibliographies, abstracts and other sources yielded 11 more. After excluding duplications, 1227 studies remained and 8 ultimately met the inclusion criteria. Compared with women undergoing non-LE primed protocols (n = 621), women exposed to LE priming (n = 468) had a lower risk of cycle cancellation [relative risk (RR): 0.60, 95% confidence interval (CI): 0.45–0.78] and an improved chance of clinical pregnancy (RR: 1.33, 95% CI: 1.02–1.72). There was no significant improvement in the number of mature oocytes obtained or number of zygotes obtained per cycle.

LIMITATIONS, REASONS FOR CAUTION: These findings are limited by the body of literature currently available. As the poor responder lacks a concrete definition, there is some heterogeneity to these results, which merits caution when applying our findings to individual patients. Furthermore, the increased clinical pregnancy rate demonstrated when using the LE protocol may be principally related to the decreased cycle cancellation rate.

WIDER IMPLICATIONS OF THE FINDINGS: The LE protocol may be of some utility in the poor responder to IVF and may increase clinical pregnancy rates in this population by improving stimulation and thereby decreasing cycle cancellation.

STUDY FUNDING/COMPETING INTERESTS: NIH K12 HD063086 (ESJ, MGT), NIH T32 HD004013-11 (KAR), F32 HD040135-10 NIH (KRO), 5K12HD000849-25 (PTJ). No competing interests.

Key words: luteal estradiol / poor responder / assisted reproduction technologies

© The Author 2013. Published by Oxford University Press on behalf of the European Society of Human Reproduction and Embryology. All rights reserved. For Permissions, please email: journals.permissions@oup.com
Introduction

A poor response to controlled ovarian hyperstimulation (COH) is frustrating, yet not uncommon in the practice of assisted reproduction technologies (ARTs). A number of studies have sought to determine the ideal COH protocol for women identified as poor responders, but most have failed to reach definitive conclusions.

One COH protocol that has shown promise in poor responders involves treatment with a short course of estradiol (E2) with or without a GnRH antagonist in the luteal phase immediately preceding COH for ART (Fanchin et al., 2003a,b; Dragisic et al., 2005; Frattarelli et al., 2008; Hill et al., 2009; Weitzman et al., 2009; Ata et al., 2011; DiLuigi et al., 2011; Ellassar et al., 2011a,b; Shastri et al., 2011; Chang et al., 2012). Proposed mechanisms include attenuation of the early luteal phase rise in FSH that poor responders often experience, and synchronization of the cohort of follicles prior to COH to improve the response to COH (DeZiegler et al., 1998; Fanchin et al., 2003a,b). Individually, many of the studies comparing ART outcomes in poor responders exposed to COH with and without luteal estradiol (LE) priming show promising results but their findings have been limited by small sample size. Accordingly, we performed a systematic review and meta-analysis to estimate the effect of LE priming prior to COH on ART outcomes in poor responders.

Methods

Literature search

Our literature search strategy was determined a priori as part of our study design following the proposal for reporting Meta-analysis of Observational Studies in Epidemiology by Stroup et al.—the MOOSE criteria (Stroup et al., 2000). Studies evaluating LE protocols and ART outcomes in poor responders were identified by searching the computerized databases MEDLINE, EMBASE and PUBMED. The searches were restricted to the English language, and were inclusive of the period up to January 2012. Search terms included variations of title terms ‘luteal OR estradiol OR poor responder AND IVF.’ Citation lists generated from the search strategy were independently screened by three authors (K.R., K.O. and P.J.) to identify relevant studies.

Titles and abstracts were manually reviewed and articles were retrieved and subjected to further evaluation if they were relevant or if there was any uncertainty as to their possible relevance. References of identified studies were subsequently manually searched for any relevant citations.

Study selection criteria

Studies were limited to those published in English. Studies were required to compare at least two treatment protocols—a ‘control’ group undergoing COH without LE and an ‘intervention’ group exposed to LE prior to COH. Cohort studies and RCTs were included. In cases in which we found that the same patient population was used in multiple publications, we only included the publication with the largest number of patients. Studies that did not specifically define their patient population as poor responders were excluded (Table I).

Data extraction

Articles meeting the inclusion criteria were kept for critical appraisal and data collection using a standard data form. Data were extracted independently by three of the authors (K.R., K.O. and P.J) and included demographic data (authors, center of study, year of publication, country, study period, number of patients included, age of patients, stimulation protocol utilized) and outcome data (clinical pregnancy, number of mature oocytes retrieved, number of zygotes, live birth, cycle cancellation). Although the definition of clinical pregnancy varied between studies, the principal definition included fetal cardiac activity as assessed by transvaginal ultrasonography after 5 weeks of gestation. Any disagreement between the three reviewers responsible for data extraction was resolved by discussion.

Methodological quality of each study was independently assessed using the criteria defined by the Downs and Black ‘Checklist for the Assessment of the Methodological Quality of Studies’ (Downs and Black, 1998). Full compliance, partial compliance or non-compliance of studies with each item on the checklist was assessed independently by two authors (K.R. and K.O.) (Table II).

Quantitative data synthesis

Data were analyzed in STATA 12.0 (College Station, TX, USA). For each outcome, raw data were entered into a 2 × 2 table and were used to calculate relative risks (RR) with 95% confidence intervals (CIs). Pooling of data was considered if there were at least two studies available for a particular...
Table II  Methodological characteristics of included studies.

| Study            | Country of origin | Study period | Study type             | Average study quality score | Number of patients (LE/control) | Patient age (years, mean ± SD) (LE/control) | Treatment protocol | Specific LE regimen | Control protocol | Primary outcome |
|------------------|-------------------|--------------|------------------------|----------------------------|--------------------------------|---------------------------------------------|---------------------|-------------------|------------------|-----------------|-----------------|
| Hill et al. (2009) | USA               | 1/2004–8/2006 | Retrospective cohort   | 17/22                      | 285 (57/228)                  | 36.3 ± 3.5 36.3 ± 3.5                  | LE with ANT stimulation | Oral micronized 17β-estradiol 2 mg twice daily | MDF              | Live birth rate |
| DiLuigi et al. (2011) | USA          | 7/2006–7/2009 | RCT                    | 18/26                      | 54 (26/28)                    | 37.5 ± 3.6 37.2 ± 3.3                  | LE/luteal ANT with ANT stimulation | One 0.1 mg transdermal E2 patch every other day | MDF              | Number of oocytes |
| Elassar et al. (2011a, b) | USA        | 1/2009–8/2010 | Retrospective cohort   | 15/22                      | 99 (52/47)                    | 39.2 ± 0.2 39.1 ± 2.8                  | LE/luteal ANT with ANT stimulation | One 0.1 mg transdermal E2 patch every other day | Letrozole/ANT     | Ongoing PR      |
| Weitzman et al. (2009) | USA          | 5/2005–4/2006 | Retrospective cohort   | 16/22                      | 121 (45/76)                   | 37.8 ± 4.7 38.6 ± 3.5                  | LE/luteal ANT with ANT stimulation | One 0.1 mg transdermal E2 patch every other day | MDF              | Cancellation rate |
| Shastri et al. (2011) | USA           | 1/2004–1/2008 | Retrospective cohort   | 16/22                      | 186 (117/69)                  | 32 ± 2.4 32 ± 2.6                     | LE/luteal ANT with ANT stimulation | One 0.1 mg transdermal E2 patch every other day | MDF              | CP              |
| Dragisic et al. (2005) | USA         | 1/2003–6/2004 | Retrospective paired cohort | 16/22                      | 132 (66/66)                   | 39.7 ± 3.5 39.7 ± 3.5                  | LE/luteal ANT with ANT stimulation | One 0.1 mg transdermal E2 patch every other day | Prior cycle       | Cancellation rate |
| Ata et al. (2011)  | Canada          | 2/2007–7/2009 | Retrospective cohort   | 18/22                      | 57 (19/38)                    | 40.0 ± 39.0                          | LE/luteal ANT with ANT stimulation | One 0.1 mg transdermal E2 patch every day    | MDF              | CP              |

LE, luteal estradiol; ANT, antagonist; MDF, microdose flare; PR, pregnancy rate; CP, clinical pregnancy; E2, estradiol; RCT, randomised control trial.
outcome. Heterogeneity between studies was tested using $I^2$ test for heterogeneity (Cochran Q) and the extent of heterogeneity was quantified using $I^2$. Pooled RR and 95% CIs were calculated for categorical outcomes (clinical pregnancy, live birth, fertilization rate and cycle cancellation) using DerSimonian and Laird random effects models. For continuous outcomes (number of oocytes retrieved per cycle, number of zygotes per cycle and number of good-quality embryos) we calculated weighted mean differences (WMDs) based on random effects models. We also performed a subgroup analysis to evaluate clinical pregnancy in studies that reported completed cycles (cycles in which patients achieved oocyte retrieval with or without embryo transfer).

Although some variation existed in stimulation regimen (Table II), results of the eight studies were pooled for our analysis as stimulation regimen alone has not been shown to improve outcomes in the poor responder (Kahraman et al., 2009).

**Outcomes**

Our primary outcome of interest was clinical pregnancy rate per cycle started. Secondary outcomes included live birth rate, number of mature oocytes retrieved, number of good-quality zygotes obtained per cycle, and rate of cycle cancellation.

**Results**

**Systematic review**

A total of 2249 publications were identified from the initial search (Fig. 1). Further search of bibliographies, abstracts and other sources revealed 11 additional studies that could possibly be included in our meta-analysis. After excluding duplications, 1227 studies remained. Individual screening

---

**Figure 1** PRISMA flow diagram for systematic review of cycle cancellation and pregnancy after luteal estradiol priming in women defined as poor responders. From Moher et al., 2009.
of each article yielded 29 studies describing the use of LE prior to COH. Upon further review of these articles, an additional 19 articles were excluded as they either did not involve poor responders to ART or they were not relevant to our study question. Although there was some inter-study variation in the definition of a poor responder (Table I), the patient population was largely similar with regard to demographics, including age. Of the remaining 10 studies, 1 was excluded owing to duplication of patient population and 1 was excluded owing to the inclusion of LE in both arms of the study, thereby eliminating the opportunity for comparison with a control arm. Ultimately, 8 studies met the final inclusion criteria giving us a total of 468 poor responders treated with LE prior to COH compared with 621 poor responders treated with COH without LE. One of the included studies was a prospective RCT and the remaining seven were retrospective cohort studies. Our quality assessment of the included studies demonstrated that they were of similar quality (Table II).

Of the included studies, six compared standard COH to LE priming along with a short course of GnRH antagonist in the luteal phase followed by COH and the remaining two studies compared standard COH to LE priming alone followed by COH.

**Reported outcomes in primary studies**

Reported outcomes included clinical pregnancy per cycle started (Hill et al., 2009; Weitzman et al., 2009; DiLuigi et al., 2011; Elassar et al., 2011a,b; Shastri et al., 2011; Chang et al., 2012), live birth (Hill et al., 2009), number of oocytes retrieved (Weitzman et al., 2009; Ata et al., 2011; DiLuigi et al., 2011; Shastri et al., 2011), number of good-quality zygotes obtained per cycle (Dragisic et al., 2005; Hill et al., 2009; Ata et al., 2011; Chang et al., 2012) and cycle cancellation (Dragisic et al., 2005; Weitzman et al., 2009; DiLuigi et al., 2011; Elassar et al., 2011a,b; Shastri et al., 2011; Chang et al., 2012).

**Meta-analysis**

Pooled RRs from random effects models revealed a significantly decreased chance of cycle cancellation for poor responders utilizing an LE protocol (RR 0.60, 95% CI 0.45–0.78) (Fig. 2A). There was no evidence of statistical heterogeneity between studies for evaluating cycle cancellation protocol as it is associated with a decrease in the risk of cycle cancellation and an increase in the chance of clinical pregnancy.

It has been proposed that LE priming may improve synchronization of the pool of follicles available to COH, thus resulting in more favorable response to COH (Fanchin et al., 2003a,b). This in turn may lead to a lower cycle cancellation rate, which is supported by our meta-analysis. Despite decreased cycle cancellation, however, LE priming did not improve the number of mature oocytes retrieved nor did it result in an increased number of zygotes in our meta-analysis. Also, when we excluded women who had the ART cycle cancelled prior to oocyte retrieval, we failed to detect a difference in the chance of clinical pregnancy in women who were primed with LE versus those who were not. This suggests that the improved chance of pregnancy among women undergoing a LE-containing stimulation protocol may be attributable to the fact that these women were more likely to make it to oocyte retrieval. In other words, they may not have had a better response to COH but rather the ART cycle was less likely to be cancelled. Because the bulk of the studies in our meta-analysis were observational, we do not know what role provider or patient bias may have played in the decision to cancel cycles in which LE priming was not employed.

**Discussion**

Poor response is a common problem in ART, and while a number of studies have investigated approaches to COH in poor responders, few have demonstrated the superiority of one COH strategy over another. Difficulties in comparing COH protocols in poor responders within and across studies may be attributed in part to the following: (i) most original research studies of ART in poor responders are small thus limiting their power to detect a significant difference between treatment protocols even if one exists and (ii) meta-analyses of the smaller studies often exclude observational studies thus ignoring the bulk of published data available for guiding the more challenging treatment of this patient population. This has indeed been the case in the literature addressing the use of LE with COH in poor responders. Given this deficiency, we performed a systematic review and meta-analysis investigating the use of LE with COH in poor responders. Our results demonstrate that women defined as poor responders benefit from the addition of LE to the stimulation protocol as it is associated with a decrease in the risk of cycle cancellation and an increase in the chance of clinical pregnancy.
In addition to the possibility that physician bias influenced the outcomes of some of the studies included in our meta-analysis, there are several other limitations to our study. We found that the definition of ‘poor responder’ varied among our included studies (Table I). We used several parameters to define poor responders for included studies. These parameters included age, basal FSH and previous poor response to stimulation. All of these criteria have been identified in the recent literature as relevant in poor response to COH and ART (Broekmans et al., 2006). Efforts have been made to create clear consensus criteria defining the patient with proven or anticipated poor ovarian response (Ferraretti et al., 2011) but applying such criteria in our approach would have left us with no studies for the meta-analysis. Variation in the definition of poor responder and variation in COH protocols and use of GnRH antagonists may explain some of the heterogeneity we
detected when comparing the number of mature oocytes retrieved and number of zygotes in women treated with LE priming prior to COH versus COH alone. If more studies investigating the effect of LE priming were available, applying meta-regression would be helpful in controlling for possible differences across studies (Colditz et al., 1994).

Most of the studies included in our meta-analysis used E2 and GnRH antagonists concurrently in the luteal phase—only two of our included studies evaluated LE alone. The biologic rationale behind utilizing LE in the poor responder is clear—the addition of E2 to the luteal phase of a cycle immediately preceding stimulation leads to synchronization of a follicle cohort and perhaps suppresses the premature rise in FSH often demonstrated in poor responders (DeZiegler et al., 1998; Fanchin et al., 2003a,b). When given in the luteal phase, GnRH antagonists likely have the same effect, thus we chose to include protocols containing both LE and LE/luteal GnRH antagonist in our treatment group analysis. We only included studies in our meta-analysis that included LE, with or without the concurrent use of GnRH antagonist. Supporting this approach, previous data have shown no difference in efficacy between the use of LE alone and the use of LE with concurrent administration of luteal antagonist (Elassar et al., 2011a,b). Also, our subgroup analysis evaluating the use of LE alone came to the same conclusion as our analysis with all studies included—that LE priming decreases the risk of cycle cancellation. In the same vein, although different treatment protocols were used to stimulate these patients during COH for IVF (i.e. flare or GnRH antagonist protocol), we chose to group these treatment protocols together in our analysis as previous work has shown no significant difference in outcomes between these stimulation regimens in the poor responder (Kahraman et al., 2009). Therefore, if a difference exists, it is likely to be secondary to the addition of modifiers, such as LE, to the basic stimulation protocol.

We went to great lengths to identify all existing studies describing our study question, but it is possible that we failed to identify all studies. We also made efforts to identify abstracts and unpublished studies investigating LE priming in poor responders, as it is possible that negative studies were not published, but we were unable to identify any when we embarked on this study. Indeed, since initiating our study, a meta-analysis on this topic was published by researchers at Nanjing Medical University, Nanjing, China (Chang and Wu, 2013). Their meta-analysis provided evidence that LE pretreatment is associated with an increased number of oocytes retrieved and decreased cycle cancellation rates in ART for poor responders but it failed to demonstrate an effect of LE pretreatment on the chance of pregnancy (Chang and Wu, 2013). While informative,
there are several important methodologic differences that differentiate Chang and Wu’s meta-analysis from ours. First, Chang and Wu (2013) followed different predetermined inclusion and exclusion criteria in their study, thus their included studies differ from ours. For example, we excluded any studies with duplicated data. Specifically, the studies by Frattarelli et al. (2008) and Hill et al. (2009) included some of the same patient data. Thus, we excluded the study by Frattarelli et al. (2008) as it was the smaller of the two studies, while Chang and Wu (2013) included both, but they did not include the studies by Ellassar et al. (2011a,b) or DiLuigi et al. (2011): both of these studies were published in 2011 and it is possible they were not in press at the time Chang and Wu performed their search. Alternately, it may be that Chang and Wu’s search terms failed to identify these studies. The differences in our included studies may have led us to find a significant effect of LE pretreatment on the chance of clinical pregnancy in poor responders, while Chang and Wu did not.

Another important difference between our methods and those of Chang and Wu (2013) is that Chang and Wu did not report whether they followed consensus criteria established for the meta-analysis of observational studies in epidemiology, e.g. the MOOSE criteria (Stroup et al., 2000). One important item in the MOOSE criteria checklist includes an assessment of quality among the included studies. In our review of study quality, we found multiple variations in the methods of the included observational studies that led us to use a random effects model to measure effect size of the LE pretreatment on all ART outcomes. Chang and Wu on the other hand chose to use a fixed-effects model to investigate some of their outcomes and a random effects model to investigate others. The fixed-effects model assumes identical treatment effects among included studies, whereas the random-effects model assumes between-study differences in treatment effects and accounts for it (Lau et al., 1992). This may have led to the differences between the two studies in reported significance and heterogeneity determined for the outcomes investigated, namely in the number of oocytes retrieved. We did not find a significant effect of LE pretreatment on quantity of mature oocytes and embryos, whereas Chang and Wu did. Also, we found significant heterogeneity among studies, whereas Chang and Wu did not. Given the methodologic variation in the studies we included and the varying definitions used for poor responders, an appropriate statistical measure of effect size accounting for heterogeneity is important. We believe the DerSimonian and Laird random-effects model to be the most effective and statistically sound model to use in this case, in the presence of such heterogeneity. The DerSimonian–Laird model takes into account the diversity often seen in patient populations and study design. This model was most appropriate for evaluating treatment protocols in the poor responder as it not only incorporates inter-study differences as well as variations in patient population in its analysis of overall treatment efficacy but also, in doing so, improves the generalizability of its results (DerSimonian and Laird, 1986).

One final limitation worth discussing further is that our meta-analysis may be limited by the fact that it largely comprises observational studies (Legro and Kunselman, 2012). As concisely stated in a recent commentary, ‘as with any meta-analysis, results are limited by flaws in the primary studies’ (Stegmann, 2012). On the other hand, we believe it is important to consider the available observational data, as this comprises the majority of the existing literature investigating the effectiveness of LE priming in poor responders. Given the limitations of meta-analysis with observational studies as discussed previously we were careful to follow MOOSE criteria in designing our study methods and reporting the findings of our data synthesis and meta-analysis (Stroup et al., 2000).

Poor responders present some of the most challenging treatment scenarios. Despite its limitations, until the results of an adequately powered, well-designed, multi-centre RCT are available on the effect of LE priming in ART, our systematic review and meta-analysis support the use of LE priming prior to COH in poor responders.

Authors’ roles

K.A.R. (corresponding author, first author) was involved in data collection and synthesis, manuscript preparation and revision, construction of figures and tables and submission of manuscript. K.R.O., P.T.J. and J.S.R. (co-authors) were involved in data collection and manuscript revision. M.G.T. (co-author) was involved in statistical analysis and served as analyst on analytical methods. E.S.J. (co-author faculty advisor) was involved in writing and revising manuscript.

Funding

This work was supported by National Institutes of Health (NIH) Grants: T32 HD0040135-11 (K.A.R.), F32 HD040130-10 (K.R.O.), SK12HD008492-25 (P.T.J.) and K12 HD063086 (E.S.J. and M.G.T.). The contents of this work are the responsibility of the authors and do not necessarily represent the official views of the NIH.

Conflict of interest

None declared.

References

Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised...


