REVIEW

Is there an association between oocyte number and embryo quality? A systematic review and meta-analysis

BIOGRAPHY

Belinda Vermey is a conjoint lecturer at the School of Women's and Children's Health at the University of New South Wales (UNSW), Australia. Belinda is also an Embryologist and is a member of Scientists in Reproductive Technologies and the Fertility Society of Australia.

Belinda G. Vermey1,*, Su Jen Chua2, Mohammad Hadi Zafarmand3, Rui Wang2, Salvatore Longobardi4, Evelyn Cottell5, Francois Beckers5, Ben W. Mol6, Christos A. Venetis1,7, Thomas D’Hooghe5

KEY MESSAGE

Considerable heterogeneity exists in the published literature on the relationship between numbers of oocytes retrieved after ovarian stimulation and live birth rates. A positive correlation was found between oocyte number and number of top/good quality embryos, suggesting that increased oocyte yield from a single stimulated cycle may maximize outcomes, although the currently available evidence has limitations.

ABSTRACT

This systematic review and meta-analysis determined the association between aspirated after ovarian stimulation and top/good quality embryos obtained in women undergoing ovarian stimulation for IVF/intracytoplasmic sperm injection (ICSI). MEDLINE, EMBASE, Scopus, CINAHL and Web of Science were searched for English-language publications on top/good-quality embryos at cleavage (day 2/3) and/or blastocyst (day 5/6) developmental stages, up to 18 November 2017. Twenty-eight studies (three prospective and 25 retrospective) reporting data on 291,752 assisted reproductive technology (ART) cycles were considered eligible. We confirmed a strong positive association between oocytes retrieved and top/good-quality day 2/3 embryos (weighted correlation coefficient $r_w = 0.791$), day 5/6 embryos ($r_w = 0.901$), metaphase II oocytes ($r_w = 0.988$), oocytes exhibiting two pronuclei ($r_w = 0.987$) and euploid embryos ($r_w = 0.851$); $P < 0.001$ for all correlations (evaluated in subsets of the 17 studies). Data from 5657 cycles showed that the group with the most oocytes aspirated had the most top/good-quality day 2/3 embryos (pooled standardized mean differences (high [>15] versus low [<4] 1.91, 95% confidence interval [CI] 1.05–2.77, $P < 0.0001$; high versus medium [4–15] 1.15, 95% CI 0.74–1.55, $P < 0.0001$; medium versus low 1.41, 95% CI 0.79–2.03, $P < 0.0001$).

Individual participant meta-analysis would enable accurate determination of these associations and other outcomes.

KEYWORDS

Embryo quality
Euploid
Oocyte number
Oocyte yield

1 School of Women's and Children's Health, University of New South Wales NSW, Australia
2 Robinson Research Institute and Adelaide Medical School, University of Adelaide SA, Australia
3 Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Public Health Research Institute, Amsterdam UMC, University of Amsterdam, The Netherlands
4 Merck Serono S.p.A, Rome, Italy
5 Merck KGaA, Darmstadt, Germany
6 Department of Obstetrics and Gynecology, Monash University, Clayton Victoria, Australia
7 Centre for Big Data Research in Health, University of New South Wales NSW, Australia

© 2019 Reproductive Healthcare Ltd. Published by Elsevier Ltd. All rights reserved.
*Corresponding author: E-mail address: belinda.vermey@outlook.com (B G.Vermey). https://doi.org/10.1016/j.rbmo.2019.06.013
INTRODUCTION

The aim of assisted reproduction technology (ART) is to help an infertile couple have a healthy baby born at term. Multifollicular growth induced by ovarian stimulation can lead to the collection of multiple oocytes, which, in turn, seems to increase pregnancy chances after ART. Hence, clinicians have for years been trying to obtain an adequate number of oocytes while minimizing the risk of ovarian hyperstimulation syndrome. Some have advocated that a higher number of collected oocytes is associated with improved outcomes (Baker et al., 2015; Macklon et al., 2006; van der Gaast et al., 2006). Others have suggested that larger oocyte yields are associated with decreased oocyte quality and hence inferior developmental potential of the resulting embryos (Pellicer et al., 1989; Valbuena et al., 2001).

Studies have shown both positive and negative associations between the number of oocytes collected and the number of good quality embryos (Kok et al., 2004). Obviously, clarifying whether oocyte yield is associated with a higher or lower number of top-quality embryos is crucial for the efficiency of the stimulated cycle. In the era of agonist triggering and vitrification, maximizing the oocyte yield could lead to higher cumulative birth rates and shorten the time to pregnancy leading to live birth. This would probably be true if it were convincingly proven that there is no negative association between the number of oocytes retrieved and the quality of the resulting embryos. Despite a number of studies (Labarta et al., 2017; Valbuena et al., 2001) that have been published on this topic, the jury seems to still be out on this.

This systematic review and meta-analysis focuses on answering the question of whether the number of oocytes aspirated after ovarian stimulation for IVF or intracytoplasmic sperm injection (ICSI) is associated with the number of top-quality embryos at the cleavage (day 2/3) and/or blastocyst stage (day 5/6) across all patient populations. The answer to this question will clarify this relationship and stimulate future studies in specific subpopulations.

MATERIALS AND METHODS

Identification of studies

This systematic review and meta-analysis was performed according to the PRISMA criteria (Liberati et al., 2009). A search strategy was developed and approved by all authors, aiming to maximize the sensitivity and specificity for identifying relevant studies (Supplementary Table 1). MEDLINE (Ovid), EMBASE (Ovid), Scopus, CINAHL (EBSCO interface) and Web of Science were searched for published studies in the English language, from inception until 18 November 2017. Reference lists of relevant publications were hand-searched for additional potentially eligible studies. Cohort studies and single arms of randomized controlled trials (treated as individual cohorts) that provided data on outcomes of interest were eligible. Case reports, case series, case-control studies and studies in which the management was modified on the basis of the expected or actual ovarian response (e.g. coating in cases of high ovarian response or split cases of IVF/ICSI) were excluded.

ART methods considered eligible were ovarian stimulation using gonadotrophins and gonadotrophin-releasing hormone analogues, cycles with one or more oocytes retrieved, and studies in which autologous fresh or frozen ejaculate spermatozoa were used for insemination. Furthermore, we included studies that provided data on the association between the number of oocytes aspirated following ovarian stimulation for IVF/ICSI and the number of metaphase II (MII) oocytes (i.e. oocytes where the extrusion of the first polar body has been completed by the time of ICSI), the number of normally fertilized oocytes (i.e. oocytes that exhibit two pronuclei during fertilization checking [e.g. at 16–20 h after insemination]) and the number of euploid embryos (i.e. with biopsied performed at the cleavage or blastocyst stage and the genetic analysis being performed by either fluorescent in-situ hybridization or comprehensive chromosomal screening).

Data extraction

Data extraction was performed independently by two authors using predefined forms. In case of missing information, the study authors were contacted to retrieve relevant data. If, within 28 days from the first e-mail, no response had been received and the missing data could not be computed or estimated (especially for the predictor and the outcomes) using the provided information, the analysis plan went ahead. Any disagreement between the two reviewers was solved by discussion with a third author.

Outcome measures

It was anticipated that the included studies would report on three forms of outcome measures.

Regression coefficient

Studies were included that reported on outcomes of interest using correlation coefficients such as Pearson product-moment correlations (correlation coefficient r) with 95% confidence intervals (CI). A ‘rule of thumb’ was used to interpret the size of a correlation coefficient (r) in which r values of <0.3, ≥0.3 to <0.5, ≥0.5 to <0.7, ≥0.7 to 0.9, and ≥0.9 to 1.0 were considered negligible, low, moderate, high and very high correlations, respectively (Hinkle et al., 2003).

Means and SD

This included studies in which the number of oocytes was expressed as different categories of response (e.g. low, medium/normal, high) and the dependent (continuous) variables were reported as either the mean or median number with SD or IQR (interquartile range). It was anticipated that the included studies would report on three forms of outcome measures.

Dealing with missing data

Authors of the eligible studies were contacted and asked to provide the coefficients of the regression model fitted to the primary and/or secondary outcome, with number of oocytes retrieved as the dependent variable. If these results were not available, the authors were asked to provide the mean and SD of primary and/or secondary outcome measures according to the following group of oocyte yield (Ferraretti et al., 2011; Polyzos and Sunkara, 2015), (i) low oocyte number: <4 oocytes retrieved, (ii) medium oocyte number: 4–15 oocytes retrieved, and (iii) high oocyte number: >15 oocytes retrieved.
Quality assessment of eligible studies
Individual studies that met the inclusion criteria were critically appraised for bias by two authors using the Quality in Prognosis Studies (QUIPS) tool (Hayden et al., 2013). In the event a disagreement could not be resolved through discussion, a third author adjudicated.

Quantitative data synthesis
Owing to the expected diversity in reporting, we considered different approaches in our predefined protocol for pooling data, including semi-partial correlation (Aloe and Becker, 2012), standardized regression coefficients (Kim, 2011), pooled correlation coefficients (Wilson, 2017) or estimation of the product-moment correlation using the Hunter and Schmidt correction method (Hunter and Schmidt, 1990) (Supplemental methods).

However, since very few studies reported a regression coefficient or correlation coefficient (or provided data to allow their calculation), a meta-analysis was not possible using these two outcome measures. Alternatively, graphical reports with available data from individual studies were produced (number of oocytes on the x-axis and number of top/good quality embryos or MII oocytes, normally fertilized oocytes or euploid embryos on the y-axis). When regression formulae were reported from individual studies, the regression line was plotted; otherwise categorical variables were used to plot the association line (e.g. by plotting the line that corresponded to the weighted average of the top/good quality embryos per average of the corresponding category of number of oocytes). This approach allowed a visual representation of the reported associations.

For studies that provided means and 95% CI (or median with IQR), pairwise pooled comparisons were performed. Estimates of effect were calculated with a random effects model to account for the expected heterogeneity and to obtain the most conservative estimates, expressed as standardized mean difference (SMD) in comparisons of groups of ovarian response. The following pairwise pooled comparisons were planned when data were available: (i) low (<4) versus medium (4–15) oocyte number; (ii) medium (4–15 oocytes) versus high (>15) oocyte number; and (iii) low (<4) versus high (>15) oocyte number.

If the original authors did not respond and/or the categorical groups of oocyte number were not the same in the original articles as the responses stated above, subjective categories of low, medium and high oocyte number were applied to calculate pairwise comparisons by calculating the mean number of oocytes obtained over the two subgroups (e.g. when a study reported two subgroups for a category [4–9 and 10–15 oocytes] in the medium oocyte number group, the weighted mean was calculated).

Assessment of heterogeneity
Between-study heterogeneity was assessed using Cochran’s Q statistic, and the I² statistic was used to estimate inconsistency in meta-analyses. The I² statistic ranges between 0% and 100%; I² = 0–25%, no heterogeneity; I² = 25–50%, moderate heterogeneity; I² = 50–75%, large heterogeneity; and I² = 75–100%, extreme heterogeneity. Significant heterogeneity was considered when P < 0.10 or I² > 50%. Attempts were made to explore the reasons for this heterogeneity (Higgins and Thompson, 2002)

The presence of publication bias was assessed using funnel plots, Egger’s test and the Begg–Mazumdar test. Statistical significance for publication bias was set at P < 0.10.

R for Windows (Version 3.4.3; R Foundation for Statistical Computing; Vienna, Austria) was used to perform statistical analyses. We used different packages for these analyses, including ‘esc’ for effect size computation for meta-analysis, ‘weights’ package for calculating weighted correlation coefficients for aggregated data, and ‘meta’, a meta-analysis package for R. Although these analyses were exploratory and descriptive, nominal statistical significance was set at P < 0.05 to guide interpretation of results.

RESULTS
The PRISMA flowchart for study identification is presented in Figure 1. The initial electronic literature search yielded 4977 publications, while hand-searching resulted in four more studies. After removing duplicates, 2565 studies were screened and 74 studies were considered eligible for full-text assessment. After further excluding 46 studies, 28 studies (published 2006–2017) reporting data on 291,752 ART cycles were considered as eligible. Individual sample sizes ranged between 40 and 256,381 ART cycles. Three of the included cohort studies were prospective studies, while the remaining 25 were retrospective. Nine of the studies included were conference abstracts only. Overall, for the outcomes of interest, the majority of studies did not provide adjustments to the analyses, and only four studies adjusted for female age (Kahraman et al., 2016; Kan et al., 2017; Kok et al., 2006; Lledo et al., 2016). Details of the methodological and population characteristics of the included studies are presented in Supplementary Tables 2, 3 and 4.

The majority of studies (86%, n = 24) categorized aspirated oocytes into groups (ranging from two to six groups), where the outcome was reported as mean (with SD or standard error [SE]), median (with or without IQR) or percentages or rates (Table 1); therefore, graphical displays of associations were made as described. Other studies reported regression coefficients when analysing associations between the continuous predictor (the number of oocytes) and the continuous primary or secondary outcomes (Kahraman et al., 2016; Kok et al., 2006; La Marca et al., 2017). One study (Hodes-Wertz et al., 2013) provided no data. In some studies, more than one statistical method was used. Eleven studies did not provide data that enabled a quantification of results, either through a pairwise comparison or through a graphical display of associations. Reasons included the data being expressed as percentages (Horton et al., 2017; Kolanyaka et al., 2012), the data not being categorized by oocyte yield (Kahraman et al., 2016; Kok et al., 2006; La Marca et al., 2017), logistic regression with aneuploidy, rather than euploidy, being the dependent variable (Colturato et al., 2010; Lledo et al., 2016), only the data on surplus embryos after a fresh embryo transfer being presented (Connell et al., 2016) and an absence of data that could be analysed (Cai et al., 2013; Hodes-Wertz et al., 2013; Milan Sanchez et al., 2014). Twelve of the 28 included studies provided data to perform pairwise comparisons (Table 1).

Owing to the extensive diversity in reporting and lack of outcome data in some studies, correlation coefficients...
could not be computed from other statistical data for each individual study as had been planned. In addition, a common effect size could not be found among the studies that had been reported directly or could have been extracted from other data provided by the individual studies. Therefore, other than performing pairwise comparisons, where feasible, graphical reports of association of outcomes were performed (17 of the 28 included studies) (TABLE 1) (Chen et al., 2015; Domingues et al., 2009; Figueira et al., 2011; Hsu et al., 2016; Ji et al., 2013; Jing et al., 2017; Kamel et al., 2014; Kan et al., 2017; Labarta et al., 2017; Li et al., 2015; Melado Vidales et al., 2017; Spitzer et al., 2015; Steward et al., 2014; Vaughan et al., 2017; Weghofer et al., 2007; Zhou et al., 2017; Zohav et al., 2017). In the current analyses, three studies categorized the number of collected oocytes and the mean of the outcome separately in different cohorts: normal and elevated progesterone (Li et al., 2015), age 21–35 years, 36–39 years and >39 years (Spitzer et al., 2015), and PCOS and controls (Weghofer et al., 2007). These cohorts were included separately in the analyses.

Quality assessment
Of the 28 included studies, risk of bias was low in 12 studies, moderate in 15 and high in one. Of note, all the included studies had a low risk of bias for study attrition, and the majority of studies had a low risk of bias for study population (19/28), measurement of prognostic factors (26/28) or measurement of outcome variables (22/28). Only a relatively small proportion of studies was scored as having a low risk of bias for the measurement and handling of confounding factors (11/28) and statistical analysis and reporting (8/28). The details of risk of bias assessments are presented in Supplementary Table 5.

Graphical reports

Number of top/good quality embryos versus number of oocytes aspirated
Fourteen studies (11 retrospective and three prospective; 16 cohorts) reported data for this comparison (FIGURE 2A) (weighted correlation coefficient (r_w) = 0.677, SE = 0.005, P < 0.001).

When comparing oocyte yield with the number of top/good quality embryos at the cleavage stage (day 2/3), the following values were obtained: r_w = 0.791, SE = 0.005, P < 0.001, n = 7 studies (five retrospective and two prospective) (FIGURE 2B).

When comparing oocyte yield with the number of top/good quality embryos at the blastocyst stage (day 5/6), values were as follows: r_w = 0.901, SE = 0.005, P < 0.001, n = 2 studies (both retrospective; four cohorts) (FIGURE 2C).

Number of oocytes aspirated versus number of MII oocytes
Five studies (two prospective and three retrospective; eight cohorts) reported data for this comparison: r_w = 0.988, SE = 0.0008, P < 0.001 (FIGURE 3A).

Number of oocytes aspirated versus number of normally fertilized oocytes
Eight studies (all retrospective; 10 cohorts) reported data for this comparison: r_w = 0.987, SE = 0.0002, P < 0.001 (FIGURE 3B).

Number of oocytes aspirated versus number of euploid embryos
Three studies (one prospective and two retrospective; four cohorts) reported data for this comparison: r_w = 0.851, SE = 0.008, P < 0.001 (FIGURE 3C).

Meta-analysis
In six studies (two prospective and four retrospective) of 5657 ART cycles (Chen et al., 2015; Hsu et al., 2016; Kamel et al., 2014; Kan et al., 2017; Labarta et al., 2017; Zhou et al., 2017), the number of top/good quality day 2/3 embryos was associated in all three pairwise comparisons with the number of oocytes aspirated: medium versus low (FIGURE 4A; pooled SMD 1.41, 95% CI 0.79–2.03; P < 0.0001; I^2 = 96%; five studies of 4169 ART cycles); high versus medium (FIGURE 4B; pooled SMD 1.15, 95% CI 0.74–1.55; P < 0.0001; I^2 = 97%; five studies of 4670 ART cycles); and high versus low (FIGURE 4C; pooled SMD 0.85, 95% CI 0.58–1.12; P < 0.0001; I^2 = 97%; five studies of 4670 ART cycles).
TABLE 1 AN OVERVIEW OF THE CONTRIBUTION OF INCLUDED STUDIES AND COHORTS TO THE GRAPHS AND ANALYSES

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of cohorts</th>
<th>Included in the graph</th>
<th>Included in the pooled pairwise comparison</th>
<th>Included in the graph</th>
<th>Included in the pooled pairwise comparison</th>
<th>Included in the graph</th>
<th>Included in the pooled pairwise comparison</th>
<th>Included in the graph</th>
<th>Included in the pooled pairwise comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hariton et al., 2017</td>
<td>1</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Kan et al., 2017</td>
<td>1</td>
<td>Overall Day 3</td>
<td>NA</td>
<td>Day 3</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>La Marca, 2017</td>
<td>1</td>
<td>Overall Day 3</td>
<td>NA</td>
<td>Day 3</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Vaughan et al., 2017</td>
<td>1</td>
<td>Overall Day 5</td>
<td>Assessment day not provided</td>
<td>Assessment day not provided - No SD for mean</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>No SD for mean</td>
<td>ND</td>
</tr>
<tr>
<td>Zhou et al., 2017</td>
<td>1</td>
<td>Overall Day 3</td>
<td>NA</td>
<td>Day 3</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Zohav et al, 2017</td>
<td>1</td>
<td>Overall Assessment day not provided</td>
<td>Assessment day not provided</td>
<td>Assessment day not provided - No SD for mean</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>No SD for mean</td>
<td>ND</td>
</tr>
<tr>
<td>Connell et al., 2016</td>
<td>1</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Hsu et al., 2016</td>
<td>1</td>
<td>Overall Day 3</td>
<td>NA</td>
<td>Day 3</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Uido et al., 2016</td>
<td>1</td>
<td>ND</td>
<td>ND</td>
<td>Day 3</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Chen et al., 2015</td>
<td>1</td>
<td>Overall Day 3</td>
<td>NA</td>
<td>Day 3</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>No SD for mean</td>
<td>ND</td>
</tr>
<tr>
<td>Li et al., 2015</td>
<td>2</td>
<td>(normal or elevated progesterone)</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>Yes</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Spitzer et al., 2015</td>
<td>3</td>
<td>(age 21–35, 36–39, &gt;39 years) Overall Day 5</td>
<td>No SD for mean</td>
<td>Yes</td>
<td>No SD for mean</td>
<td>Yes</td>
<td>No SD for mean</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Cil et al., 2014/Kahraman et al., 2016</td>
<td>1</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
</tr>
<tr>
<td>Milan Sanchez et al., 2014</td>
<td>1</td>
<td>Overall Assessment day not provided</td>
<td>Assessment day not provided</td>
<td>Assessment day not provided - No SD for mean</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>Sample sizes per categories were not reported</td>
<td>ND</td>
</tr>
<tr>
<td>Steward et al., 2014</td>
<td>1</td>
<td>Overall Assessment day not provided</td>
<td>Assessment day not provided</td>
<td>Assessment day not provided - No SD for mean</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>Sample sizes per categories were not reported</td>
<td>ND</td>
</tr>
<tr>
<td>Cai et al., 2013</td>
<td>1</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Hodes-Wertz et al., 2013</td>
<td>1</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
</tr>
<tr>
<td>Ji et al., 2013</td>
<td>1</td>
<td>Mean and SD of mean were not reported</td>
<td>Mean and SD of mean were not reported</td>
<td>Mean and SD of mean were not reported</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>No SD for mean</td>
<td>ND</td>
</tr>
<tr>
<td>Kolankaya et al., 2012</td>
<td>1</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Figueira et al., 2011</td>
<td>1</td>
<td>Overall Day 3</td>
<td>NA</td>
<td>No SD for mean</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No SD for mean</td>
<td>ND</td>
</tr>
</tbody>
</table>

(continued on next page)
<table>
<thead>
<tr>
<th>Study</th>
<th>No. of cohorts</th>
<th>Included in the graph</th>
<th>Included in the pooled pairwise comparison</th>
<th>Included in the graph</th>
<th>Included in the pooled pairwise comparison</th>
<th>Included in the graph</th>
<th>Included in the pooled pairwise comparison</th>
<th>Included in the graph</th>
<th>Included in the pooled pairwise comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colturato et al., 2010</td>
<td>1</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Dominguez et al., 2009</td>
<td>1</td>
<td>Overall</td>
<td>Assessment day not provided</td>
<td>Assessment day not provided</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Weghofer et al., 2007</td>
<td>2 (PCOS or controls)</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Kok et al., 2006</td>
<td>1</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
<td>No categories</td>
</tr>
<tr>
<td>Labarta et al., 2017</td>
<td>1</td>
<td>Overall</td>
<td>Day 3</td>
<td>NA</td>
<td>Day 3</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Jing et al., 2017</td>
<td>1</td>
<td>Overall</td>
<td>Assessment day not provided</td>
<td>Assessment day not provided</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Karra et al., 2014</td>
<td>1</td>
<td>Overall</td>
<td>Day 3</td>
<td>NA</td>
<td>Day 3</td>
<td>Yes</td>
<td>Yes</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Melado Vidales et al., 2017</td>
<td>1</td>
<td>Overall</td>
<td>Assessment day not provided</td>
<td>Assessment day not provided</td>
<td>Yes</td>
<td>Yes</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Total cohorts included</td>
<td>32</td>
<td>16</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>5</td>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>

NA, not available; ND, not determined; PCOS, polycystic ovary syndrome.
1.91, 95% CI 1.05–2.77; P < 0.0001 (I² = 98%, four studies of 2475 ART cycles)). Only one study (Vaughan et al., 2017) reported on the number of top/good quality day 5/6 embryos for each of the categories; meta-analysis was therefore not possible.

Four studies (two prospective and two retrospective; 4332 ART cycles) reported on the association between the number of oocytes aspirated and the number of MII embryos (Figueira et al., 2011; Kamel et al., 2014; Li et al., 2015; Melado Vidales et al., 2017): medium versus low (Supplemental Figure 1A; pooled SMD 1.99, 95% CI 1.85–2.13; P < 0.0001 [I² = 15%, three studies, four cohorts of 3356 ART cycles]); high versus medium (Supplemental Figure 1B; pooled SMD 1.89, 95% CI 1.35–2.43; P < 0.0001 [I² = 92%, three studies, four cohorts of 3250 ART cycles]); and high versus low (Supplemental Figure 1C, pooled SMD 3.63, 95% CI 2.64–4.61; P < 0.0001 [I² = 96%, two studies, three cohorts of 2058 ART cycles]).

Two studies (both retrospective; 785 ART cycles) reported on the association between the number of oocytes aspirated and the number of normally fertilized embryos (Jing et al., 2017; Kan et al., 2017): medium versus low (Supplemental Figure 2A; pooled SMD 1.75, 95% CI 0.95–2.54; P < 0.0001 [I² = 73%; two studies of 582 ART cycles]); high versus medium (Supplemental Figure 2B; pooled SMD 1.89, 95% CI 1.70–2.08; P < 0.0001 [I² = 0%; two studies of 715 ART cycles]); and high versus low (Supplemental Figure 2C; pooled SMD 3.19, 95% CI 1.97–4.40; P < 0.0001 [I² = 78%; two studies of 273 ART cycles]).

**FIGURE 2** The association between the number of oocytes aspirated and (A) the number of top/good quality embryos ($r_w = 0.677$, SE = 0.005, $P < 0.001$), (B) the number of top/good quality embryos at the cleavage stage (day 2/3) ($r_w = 0.791$, SE = 0.005, $P < 0.001$), and (C) the number of top/good quality embryos at the blastocyst stage (day 5/6) ($r_w = 0.901$, SE = 0.005, $P < 0.001$). $r_w$ weighted correlation coefficient; SE, standard error.
Three studies (one prospective and two retrospective) reported on the number of oocytes aspirated and the number of euploid embryos (Kan et al., 2017; Labarta et al., 2017; Weghofer et al., 2007): medium versus low (FIGURE 5A; pooled SMD 0.60, 95% CI 0.30–0.89, \( P < 0.0001 \) [one study of 644 ART cycles; Kan et al., 2017]); high versus medium (FIGURE 5B; pooled SMD 0.70, 95% CI 0.55–0.85, \( P < 0.0001 \) \( [r^2 = 0\% \), four studies of 875 ART cycles]); and high versus low (FIGURE 5C; pooled SMD 1.00, 95% CI 0.67–1.32, \( P < 0.0001 \) [one study of 329 ART cycles; Kan et al., 2017]).

**DISCUSSION**

The results of this systematic review and meta-analysis suggest a positive association between the number of top/good quality embryos and the number of oocytes retrieved after IVF or ICSI. It also suggests an association between oocyte yield and MII and normally fertilized oocytes and euploid embryos. In the pooled pairwise comparisons, the high (>15) oocyte number group was associated with more top/good quality embryos, MII oocytes, normally fertilized oocytes and euploid embryos than the medium (4–15) and low (<4) oocyte number groups; the medium group was associated with better outcomes than the low oocyte number group. These findings were confirmed visually when plotting the average number of oocytes aspirated with the average number of top/good quality cleavage and/or blastocyst embryos, MII oocytes, normally fertilized oocytes and euploid embryos with weighted correlation coefficients with large effect sizes for all outcomes.

To date, there have been no published meta-analyses evaluating the relationship between the number of oocytes aspirated following ovarian stimulation for IVF/ICSI and the number of top/good quality embryos. Two meta-analyses evaluated the oocyte yield in relation to pregnancy rates (van Loendersloot et al., 2010; Verberg et al., 2009). One meta-analysis (Verberg et al., 2009) made a distinction between conventional stimulation and mild stimulation. In patients treated with mild ovarian stimulation, the ongoing pregnancy rate was positively correlated with an increasing number of oocytes (up to 4–6 oocytes) and declined if seven or more oocytes were obtained. In patients treated with conventional ovarian stimulation, the ongoing pregnancy rate was positively correlated with the number of oocytes retrieved up to 10–12 oocytes, and declined if 13 oocytes or more

![Figure 3](image-url)
were retrieved. Both meta-analyses (van Loendersloot et al., 2010; Verberg et al., 2009) demonstrated a clear and positive association between an increasing number of oocytes retrieved and the probability of pregnancy after IVF. Many other investigators (Drakopoulos et al., 2016; Ji et al., 2013; Magnusson et al., 2018; Polyzos and Sunkara, 2015; Sunkara et al., 2011; Toftager et al., 2017; Vaughan et al., 2017) have confirmed a strong positive correlation between the number of oocytes retrieved and live birth rates, and between the number of oocytes retrieved and the cumulative live birth rate per started cycle.

Despite the recent and robust evidence showing a positive correlation between the number of oocytes retrieved, live birth rates and cumulative live birth rates, older studies have proposed that embryo quality and the likelihood of a clinical pregnancy decrease as the number of aspirated oocytes increases following ovarian stimulation in an IVF/ICSI cycle (Pellicer et al., 1989; Valbuena et al., 2001). One group suggested a detrimental effect at the chromosome level, with a decreased rate of euploidy as oocyte yield increased (Baart et al., 2007). Possible contributors to this association include the effect of supraphysiological oestradiol concentrations on the developing oocyte and the potential negative effects of oestradiol, and especially progesterone, on endometrial receptivity (Venetis et al., 2013; Venetis et al., 2015). Another group (Racca et al., 2018) observed a positive association between elevated serum progesterone concentrations at the end of the follicular phase and an increased

**FIGURE 4** The number of oocytes aspirated versus number or top/good quality cleavage (day 2/3) embryos. (A) Medium oocyte number versus low oocyte number, (B) high oocyte number versus medium oocyte number, and (C) high oocyte number versus low oocyte number.
number of oocytes retrieved. However, the cumulative live birth rates per cycle started were significantly higher (36% and 35% versus 24%) in the subgroups with normal or elevated progesterone compared with the low-progesterone subgroup. These two subgroups also had the highest median number of oocytes (seven and 10) compared with the subgroup with low progesterone (five). This suggests that the number of good quality oocytes and embryos was more important for determining reproductive outcomes than the rise in serum progesterone. Other recent studies have confirmed that there is a positive correlation between the number of oocytes retrieved and the number of good quality and euploid embryos, therefore ruling out a detrimental effect of increased ovarian response on embryo quality or rate of euploidy (Ata et al., 2012; Kan et al., 2017; Labarta et al., 2017).

Another theory that has been suggested is the impact of ovarian stimulation on rates of normal fertilization. A recent study reported a decreased rate of normal fertilization in patients who responded with high and moderate oocyte yields compared with patients who responded with low oocyte yields (Hsu et al., 2016). Although in that study an increased number of oocytes retrieved (higher responders) was accompanied by lower fertilization rates (owing to the increased number of immature oocytes), there were still more fertilized oocytes in the high responders compared with the low responders. The current systematic review and meta-analysis shows a strong and positive association between the number of oocytes retrieved and the number of fertilized oocytes.

Multiple statistical approaches were used to analyse the available evidence. Although the different statistical approaches resulted in consistent conclusions, the large heterogeneity and low number of studies, particularly for secondary outcomes, require these results to be interpreted with caution. More specifically, although the direction of the effect seems to be consistent, a calculation of an effect size from all available studies is not feasible.

This systematic review has some limitations that need to be discussed. First, it is fully recognized that low,
medium and high oocyte number may be a consequence of low, normal or high ovarian response. However, ovarian response is determined by many factors (e.g. patient baseline characteristics, type and starting dose of gonadotrophins). It was not the aim of this paper to correlate oocyte number with ovarian response, but rather oocyte number with good quality embryo number, recognizing that oocyte number can be determined by several factors, including ovarian response. The analysis is based on observational studies, most of which were retrospective and of moderate quality, therefore, the presence of confounding bias cannot be excluded. Second, the studies included exhibited heterogeneity in terms of clinical protocols, outcome reporting strategies, research designs and statistical analyses. Because of the low number of included studies in the pairwise comparisons, an evaluation of the sources of heterogeneity, through either a meta-regression analysis or a subgroup analysis, could not be performed. Additionally, a number of the included studies categorized the number of retrieved oocytes in multiple categories (e.g. low oocyte number, medium oocyte number and high oocyte number). Although this approach might be clinically intuitive, it generally reduced the power of the analysis and also led to problems with interpretation, since the cut-offs varied substantially between studies. The additional graphical reports of association of outcomes presented when a pairwise comparison could not be performed did not alter the conclusions drawn from the results of the meta-analysis. Finally, many of the studies included did not report results after adjusting for known confounding factors that might explain the heterogeneity (e.g. age). Given the substantial heterogeneity of the eligible studies and, although a random effects model was used to obtain the most conservative estimates, the results should be interpreted with caution.

On the basis of the results of this systematic review and meta-analysis, increased numbers of top/good quality embryos at cleavage (day 2/3) and/or blastocyst (day 5/6) stage were associated with an increased number of oocytes aspirated after ovarian stimulation. As the use of agonist triggering and vitrification increases worldwide, with promising results on cumulative pregnancy rates, clinicians should expect that a larger oocyte yield will result in a larger number of top/good quality embryos, which can potentially maximize cumulative live birth rates or increase the number of live births from a single stimulated cycle and shorten the time to pregnancy leading to live birth. However, before such a strategy can be formally recommended, data should be collected through properly controlled studies. It should be stressed that the association reported here is likely to be driven by patient characteristics. A better ovarian response is probably a reflection of a better prognostic profile, with respect to oocyte quantity and quality, but this could not be investigated in this systematic review. At the level of the individual woman, slight fluctuations in oocyte level did not seem to lead to increased pregnancy rates in several trials, and increased gonadotrophin dose in women who are expected to have a poor response leads to more oocytes being recovered and more embryos transferred but not in more babies born (Arce et al., 2014; Baart et al., 2007; Nyboe Andersen et al., 2017; Sterrenburg et al., 2011; van Tilborg et al., 2017).

This systematic review and meta-analysis identified differences in research design when evaluating this particular research question. Owing to these differences and the lack of reporting of certain information, appropriate adjustment for individual patient characteristics could not be performed, but this could be overcome in future analyses by using individual participant data. There are several advantages of performing an individual participant data meta-analysis for this research question. These include an overall estimation of the effect, a better analysis of subgroup data (e.g. whether frozen or fresh spermatozoa were used), the ability to combine different scales of measurement, the ability to analyse outcome measures that are defined differently, an in-depth exploration of patient factors, exploration of the effect of patient exclusion and for combating poor reporting (Broeze et al., 2009; Broeze et al., 2010). Furthermore, future individual studies evaluating this research question should use advanced regression methods while controlling for major confounders.

**CONCLUSIONS**

This systematic review and meta-analysis indicates that the number of top/good quality embryos would increase with increasing oocyte yield. More investigation is required to ascertain the optimum level of stimulation that will give an increased embryo yield but does not compromise patient convenience and safety owing to over-stimulation. Owing to the substantial heterogeneity among the available studies, an individual participant data meta-analysis is needed to accurately determine the functional form of the association between the numbers of oocytes retrieved and the number of top/good quality embryos with appropriate adjustment for individual patient characteristics.

**ACKNOWLEDGEMENTS**

The investigators would like to thank the following authors and their contributors for providing us with the extra information of the included studies: Hariton et al., 2017; Kahraman et al., 2016; Kan et al., 2017; Kok et al., 2006; Zhou et al., 2017. We would also like to thank the anonymous peer reviewers who provided constructive feedback on the first version of the manuscript. Editorial support was provided by Steven Goodrick of InScience Communications, London, UK. This project was supported and funded by Merck KGaA.

**SUPPLEMENTARY MATERIALS**

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.rbmo.2019.06.013.


Hunter, J.E., Schmidt, F.L. 1990 Methods of Meta-Analysis: Correcting Error and Bias in Research Findings. SAGENewbury Park, CA


La Marca, A., Minasi, M.G., Sghinolfi, G., Greco, P., Argento, C., Grisendi, V., Fiorentino, F., Greco, E. Female age, serum antimullerian hormone level, and number of oocytes affect the rate and number of euploid blastocysts in in vitro fertilization/intracytoplasmic sperm injection cycles. Fertil. Steril. 2017; 108: 777–783


Milan Sanchez, M., Rodrigo, L., Mateu, E., Penado, V., Garcia Herrera, S., Campos Galindo, I., Poo, M.E., Simon, C., Rubio, C. What is the minimal number of MII oocytes or day-3 embryos for a successful comprehensivechromosomal Screening (CCS) cycle. Hum. Reprod. 2014; 29: i301


Pellicer, A., Ruiz, A., Castellvi, R.M., Calatayud, C., Ruiz, M., Tarr, J.J., Miró, F. How is the retrieval of high numbers of oocytes desirable in patients treated with

REFERENCES
gonadotrophin-releasing hormone analogues (GnRHAs) and gonadotrophins? Hum. Reprod. 1989; 4: 536–540
Valbuena, D., Martin, J., de Pablo, J.L., Remohi, J., Pellicer, A., Simon, C. Increasing levels of estradiol are deleterious to embryonic implantation because they directly affect the embryo. Fertil. Steril. 2001; 76: 962–968
Received 1 April 2019; received in revised form 24 June 2019, accepted 28 June 2019.