Effectiveness of tonifying-kidney and regulating-liver therapy on diminished ovarian reserve: a systematic review and Meta-analysis of randomized controlled trials

Liu Liuqing, Liu Yanfeng, Yang Ming, Xu Guiqin, Li Ruiqi, Xu Xiuli, Pan Xue, Liang Jialing

**METHODS:** The literature was comprehensively searched up to August 2019 using four Chinese and three English electronic databases to extract randomized clinical trials (RCTs) comparing Traditional Chinese Medicine tonifying-kidney and regulating-liver prescriptions (combined with hormone therapy or not) with Western Medicine. Data quality evaluation was conducted using the Cochrane risk of bias tool. Meta-analysis was conducted using Revman 5.3 software with effect estimates presented as mean difference (MD), risk ratio (RR), and 95% confidence interval (CI).

**RESULTS:** A total of nine RCTs with 512 participants were extracted and eligible for Meta-analysis. There were no significant differences between Chinese medicine and Western Medicine on basal serum follicle-stimulating hormone (FSH) level (MD 0.11, 95% CI = 0.52 to 0.74, 392 participants, seven trials), anti-Müllerian hormone level (MD 0.48, 95% CI = 0.62 to 1.58, 95 participants, two trials), and the FSH and luteinizing hormone ratio (MD 0.01, 95% CI = 0.95 to 0.96, 115 participants, two trials). Chinese medicine was more effective at improving Traditional Chinese Medicine symptom scores (TCMSS) (MD = 2.39, 95% CI = 3.83 to –0.94, 160 participants, three trials), effective rate of TCMSS (RR 1.18, 95% CI 1.02 to 1.36, 160 participants, three trials), antral follicle count (AFC) (MD 0.55, 95% CI 0.05 to 1.04, 155 participants, three trials), and FSH levels at 3 months post-treatment (MD = 4.77, 95% CI = 6.09 to –3.45, 137 participants, two trials).

**CONCLUSION:** Compared with Western Medicine, tonifying-kidney and regulating-liver therapy is more effective at relieving symptoms and improving AFC and FSH at 3 months post-treatment.
INTRODUCTION

Diminished ovarian reserve (DOR) is a normal physiologic occurrence for older women because of the age-related decline in ovarian function. However, many women experience DOR much earlier. This is termed pathologic DOR, and describes women of reproductive age with normal menstruation who have a more diminished response to ovarian stimulation or fertility than comparable women.

Pathologic DOR may lead to decreased fertility and poor pregnancy outcomes, such as high rates of pregnancy loss. Abnormal ovarian reserve is also closely related to poor ovarian response, according to the Bologna criteria published by the European Society of Human Reproduction and Embryology, and may be the prodromal stage of ovarian failure. DOR can also adversely affect the outcome of in vitro fertilization-embryo transfer (IVF-ET) to some extent. According to 1995-1998 data from the United States, 10.55% of patients (1034/9802) in one infertility clinic were diagnosed with DOR. Another study demonstrated that DOR prevalence increased from 19% to 26% from 2004 to 2011 among 181 536 fresh, autologous assistive reproductive technology (ART) cycles reported to the Society for Assisted Reproductive Technology Clinic Outcomes Reporting System. Thus, early medical intervention for DOR is urgently needed to reverse declining ovarian function and promote natural pregnancy or increase the IVF-ET pregnancy rate.

In 2015, the Practice Committee of the American Society for Reproductive Medicine noted the lack of a "uniformly accepted definition of DOR," which makes the criteria for DOR difficult to determine. DOR is currently clinically diagnosed by an abnormal ovarian reserve test comprising elevated but not menopausal basal follicle-stimulating hormone (FSH), low anti-Müllerian hormone (AMH), low antral follicle count (AFC), and other less frequently used markers (like FSH/LH [luteinizing hormone] ratio, estradiol level, and previous IVF outcome) among women with regular menstrual periods.

DOR can be induced by multiple causes, which remain obscure. In addition to some idiopathic causes, impaired ovarian reserve may be associated with genetic factors, autoimmune factors, iatrogenic factors (such as radiotherapy, chemotherapy, or surgery), and psychological and environmental factors. Underlying mechanisms include excessive follicle atresia, ovulation dysfunction, direct destruction of organs or tissues, and disturbances in the hypothalamic-pituitary-ovarian axis.

Age is another risk factor, whereas late menarche (menarche age > 13 years) is a protective factor for the decline of ovarian reserve function.

There is currently no definitive and effective method to restore ovarian function in patients with DOR. In addition to lifestyle changes and the avoidance of harmful factors, hormonal replacement therapy (HRT), which includes sequential therapy with estrogen and progesterone (artificial menstrual cycle) and combined estrogen/progesterone, is a primary treatment for DOR patients to correct low estrogen status, prevent the reproductive organs from shrinking, maintain regular menstruation, and prevention related complications. The androgen therapy of dehydroepiandrosterone supplementation can also improve ovarian function, lower miscarriage rate, reduce aneuploidy, and increase pregnancy chances. Various ovulation induction treatments and assisted reproductive techniques are also available for infertile DOR patients who wish to become pregnant. However, although it is fast and effective, HRT has some contraindications. Assisted reproductive technology also has a high risk of failure and is expensive. The recovery of ovarian function and the treatment of infertility caused by DOR therefore remains a substantial challenge.

However, Traditional Chinese Medicine (TCM) has a good therapeutic effect on DOR. TCM has an overall effect on regulation and restoration of ovarian function, as it can regulate endocrinological factors and protect ovarian function from damage. Meta-analysis indicates that TCM may provide an effective and safe alternative therapy for patients with DOR. From a TCM perspective, DOR is categorized under menopause, metrorrhagia and metrotaxis, infertility, etc. The most basic type of DOR is the TCM kidney deficiency pattern. The substantial mental stress that arises from the important social roles women play in modern society means that liver depression is a common pattern. Some research indicates that kidney deficiency and liver depression is a common pattern, so tonifying-kidney and regulating-liver therapy is frequently used. TCM prescriptions for this therapy consist mainly of kidney-tonifying medicine (which nourishes kidney Yin and kidney essence, and reinforces kidney Yang and kidney Qi) and liver-regulating medicine (which smooths liver Qi, pacifies liver Yang, emolliates the liver, and clears liver heat). These medicines replenish the essence and blood and regulate the movement of Qi and blood, which restores the reproductive function of DOR patients. In this study, we aimed to evaluate the role that kidney-tonifying and liver-regulating therapy plays in DOR treatment using a systematic review and Meta-analysis.

METHODS

Inclusion criteria

Included studies had to meet four criteria. First, only
randomized controlled trials (RCTs) reporting the effectiveness of tonifying-kidney and regulating-liver therapy were eligible for inclusion (regardless of blinding). Second, eligible participants had to meet the DOR criteria. As mentioned above, there are currently no uniform diagnostic criteria for DOR. According to the Bologna criteria, AMH and AFC are two important indicators of ovarian reserve. One study of DOR diagnostic criteria based on a PubMed search summarized 14 studies that defined DOR and showed that serum FSH, AFC, and AMH levels were the most frequently evaluated indicators, and were thus most commonly used in diagnosis. On the basis of the literature and these criteria, combined with the definition of DOR in the Chinese Expert Consensus about the Diagnosis And Treatment of Premature Ovarian Insufficiency, we defined DOR as an abnormal ovarian reserve test result in patients aged <40 years: AFC (monitored by ultrasonography) < 5, or AMH < 1.1 ng/mL, or 10 IU/L < basal FSH < 40 IU/L, or basal FSH/basal LH > 3 (basal FSH and basal LH should be tested on menstruation cycle days 2-4). Third, regarding interventions, the experimental group had to be treated with an oral TCM decoction comprising mainly kidney-tonifying drugs (including kidney Yin-nourishing drugs, kidney essence-nourishing drugs, kidney Yang-reinforcing drugs, and kidney Qi-reinforcing drugs) and liver-regulating drugs (liver-smoothing drugs, liver-emptiling drugs, liver-pacifying drugs, and liver heat-clearing drugs) according to the Pharmacopoeia of the People’s Republic of China (2015) and the Traditional Chinese Pharmacy textbook, combined or not with common Western Medicine (WM) treatments (the presence of herbs with other efficacies in prescription or modified prescriptions were allowed). The control group had to be treated with HRT, androgen therapy, ovulation induction treatment, or ART. Finally, included studies had to report the results of FSH, or AFC, or AMH, which we designated as the main outcome indicators owing to the important role they play in the diagnosis of DOR.

**Exclusion criteria**

Studies were excluded if they met the following conditions: (a) the sum number of kidney-tonifying drugs and liver-regulating drugs was less than 60% of the total number of drugs in the TCM prescription (including drugs added circumstantially), or the number of kidney-tonifying medicines was less than two, or the number of liver-regulating drugs was less than two; (b) the therapeutic regimen details were not clearly reported; (c) FSH, AMH, or AFC results were not included in the outcome indicators; (d) data were incorrect or prescriptions were incomplete.

**Outcome measurements**

Following the commonly used test indexes reported in the literature, the main outcome indicators we selected were basal serum FSH levels, AMH, and AFC. TCM symptom score (TCMSS), effective rate of TCMSS, FSH/LH ratio, and basal serum FSH tested at 3 months post-treatment were selected as additional outcome indicators.

**Data sources and search strategies**

Two authors (Liu LQ and Li RQ) were in charge of the data search. An extensive literature search was conducted to identify potentially eligible studies in the following electronic databases: China National Knowledge Infrastructure (CNKI), WanFang Database, Chinese VIP Information, SinoMed, PubMed, Cochrane Library, and EMBASE. The search strategy used the following terms: “diminished ovarian reserve” and “(Traditional Chinese Medicine) or (Chinese medicine) or (herbs) or (complementary or alternative medicine)” and “randomized.” Records were searched up to August 2019 and no language restriction was applied. For instance, the search strategy employed for PubMed is shown below.

#1 diminished ovarian reserve [Title/Abstract] OR diminished ovarian reserve [MeSH Terms]
#2 decreased ovarian reserve [Title/Abstract] OR decreased ovarian reserve [MeSH Terms]
#3 ovarian dysfunction [Title/Abstract] OR ovarian dysfunction [MeSH Terms]
#4 #1 OR #2 OR #3
#5 Traditional Chinese Medicine [Title/Abstract] OR Traditional Chinese Medicine [MeSH Terms]
#6 Chinese medicine [Title/Abstract] OR Chinese medicine [MeSH Terms]
#7 herbs [Title/Abstract] OR herbs [MeSH Terms]
#8 complementary or alternative medicine [Title/Abstract] OR complementary or alternative medicine [MeSH Terms]
#9 #5 OR #6 OR#7 OR #8
#10 randomized [All fields]
#11 #4 AND #9 AND #10

**Selection of studies**

Two reviewers (Liu Liqing and Pan Xue) selected studies independently. Records obtained were first checked to eliminate duplicates. Then preliminary screening was conducted by reading the title and abstract to exclude studies unrelated to the clinical treatment of DOR by TCM. The full text of the retained records was then downloaded and read to select eligible studies according to the inclusion and exclusion criteria. If the full text of a study could not be obtained, we contacted the author to request the relevant content. If the author did not reply or did not provide the requested content, the study was excluded. Any disagreement in this process was resolved with assistance from a third party (Liu YF).

**Data extraction**

Two authors (Liu Liqing and Xu Xiuli) screened the included studies independently and extracted relevant data.
RESULTS

Assessment of risk of bias in included studies

Two reviewers (Liu Liuqing and Yang Ming) independently assessed the quality of included studies using the Cochrane Handbook risk of bias assessment tool. This tool comprises seven aspects: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting data, and other bias. Each potential source of bias was graded as high, low, or unclear. Any disagreement was resolved by discussion, with assistance from a third party (Yang M) if necessary.

Statistical analysis

The Meta-analysis was conducted using Revman 5.3 (Copenhagen, DK, the Nordic Cochrane Centre, Cochrane Collaboration, 2014) to combine more than one trial to estimate pooled intervention effects. Dichotomous data were summarized using risk ratios (RRs) with 95% confidence intervals (95% CI). Continuous data were expressed as mean differences (MDs) with 95% CI. Statistical heterogeneity was tested using the $\chi^2$ test and examining $I^2$, and data were combined in a fixed-effect model if statistical heterogeneity was low. When $P < 0.1$ or $I^2 > 50\%$, which indicated the possibility of statistical heterogeneity, a random-effect model was used. In the event that a meta-analysis was not possible, we planned to use a narrative analysis of the results from individual studies. Subgroup analysis was conducted to determine the evidence for a different intervention effect (only TCM treatment or TCM treatment combined with WM). A funnel plot was used to detect publication bias.

A protocol of study methods was registered in PROSPERO and published on the website (https://www.crd.york.ac.uk/prospero/). Trial registration number in PROSPERO: CRD42019120832.

Risk of bias in included studies

We evaluated the included studies from seven aspects (Figure 2). All were RCTs, but only three (23, 26, 27) reported the randomization procedure (use of a random number table). Therefore, these three studies were classed as "low risk of bias" and the other six studies classed as "unclear risk of bias." None of the nine studies mentioned allocation concealment so they were classed as having an "unclear risk of bias." None of the included studies mentioned blinding of participants and personnel. In consideration of the obvious difference in dosage forms between the two patient interventions, we classed all the studies as "high risk of bias" for this aspect. None of the studies mentioned blinding of assessment, so all were classed as "unclear risk of bias" for this aspect. Regarding incomplete outcome data, one study (23) was classed as "high risk of bias" because of the differences in sample size between different outcomes without explanation. Six studies were classed as "unclear risk of bias" because of the inconsistent number of participants before and after grouping (which did not significantly influence the data analysis) or the absence of loss to follow-up reporting. Another two studies (23, 27) were classed as "low risk of bias." Regarding selective reporting, one study (23) did not report the outcomes to be observed in the methods, so we had to class it as "unclear risk of bias." The other eight studies were classed as "low risk of bias" for this aspect, as the outcomes to be observed that were mentioned in the methods were all reported in the results. Regarding other forms of bias, six studies (23, 24, 26-28) were academic dissertations and one study (23) was supported by the Shanghai Municipal Health Bureau, so they were classed as "low risk of bias." The other two studies were journal articles (25, 26) with no declarations of interest or support and were assessed as "unclear risk of bias."
Figure 1 Flow chart of study selection
CNKI: China National Knowledge Infrastructure; VIP: China Science and Technology Journal Database; EMBASE: Excerpta Medica Database.

**Effects of interventions: FSH**
All of the included studies examined FSH, and a meta-analysis could be conducted from data from eight studies. Of these, one tested TCM combined with WM, and the others tested TCM versus WM. These seven studies showed no heterogeneity ($P = 0.70 > 0.1$, $I^2 = 0\% < 50\%$), so the fixed-effect model was applied. The results showed no obvious difference between TCM and WM in improving FSH levels for patients with DOR ($MD = 0.11$, 95% CI $-0.52$ to $0.74$, 392 participants, seven trials), and no statistically significant difference ($P = 0.73 > 0.05$). For the study comparing TCM combined with WM versus WM, the results ($MD = -1.87$, 95% CI $-3.68$ to $-0.06$, 60 participants, one trial) showed that the effectiveness of integrated therapy may be better than only WM ($P = 0.04 < 0.05$). A meta-analysis of these eight studies showed no significant difference in FSH levels of the experimental group and control group ($-0.11$, 95% CI $-0.70$ to $0.49$, 452 participants, eight trials) (Figure 3).

The remaining study, which was not included in the Meta-analysis, divided participants into subgroups according to whether FSH levels were normal ($<10$ mIU/mL) or not ($10-40$ mIU/mL) before treatment and observed changes in FSH levels after treatment in different subgroups. The baselines (FSH levels before treatment) of the TCM group and WM group in two subgroups were both consistent. In the normal FSH subgroup, FSH level after treatment was ($16.74 \pm 7.50$) mIU/mL in the TCM group ($n = 19$) and ($14.74 \pm 4.59$) mIU/mL in the WM group ($n = 19$). There was no significant difference between the groups ($P = 0.32 > 0.05$). In the normal FSH subgroup, FSH level after treatment was ($8.68 \pm 0.71$) mIU/mL in the TCM group ($n = 11$) and ($8.61 \pm 0.76$) mIU/mL in the WM group ($n = 11$) and there was no significant difference between the groups ($P = 0.82 > 0.05$).

**Effects of interventions: AMH**
AMH was examined in two studies. These showed heterogeneity ($P = 0.001 < 0.1$, $I^2 = 91\% > 50\%$), so the random-effect model was used. The results indicat-
**Table 1 Characteristics of included studies**

<table>
<thead>
<tr>
<th>Study</th>
<th>Sex</th>
<th>Age (T/C)</th>
<th>Sample size</th>
<th>Intervention</th>
<th>Control</th>
<th>Outcome measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cai YM et al. 2017, 2018</td>
<td>Female</td>
<td>33.77±4.01 / 34.17±4.21</td>
<td>30 / 30</td>
<td>Artificial menstrual cycle (estradiol valerate / medroxyprogesterone)</td>
<td>Chinese medicine</td>
<td>TCMS; pregnancy rate, FSH, LH, FSH/LH, AMH, OV, AFC, AE</td>
</tr>
<tr>
<td>Guan Y et al. 2016</td>
<td>Female</td>
<td>35/35</td>
<td>29 / 30</td>
<td>Kidney deficiency and liver depression</td>
<td>Chinese medicine</td>
<td>TCMS; pregnancy rate, FSH, LH, FSH/LH, AMH, OV, AFC, AE</td>
</tr>
<tr>
<td>Li SP et al. 2004</td>
<td>Female</td>
<td>20±20</td>
<td>30 / 25</td>
<td>Kidney deficiency and liver depression</td>
<td>Chinese medicine</td>
<td>TCMS; pregnancy rate, FSH, LH, FSH/LH, AMH, OV, AFC, AE</td>
</tr>
<tr>
<td>Lu YJ et al. 2016</td>
<td>Female</td>
<td>30±30</td>
<td>24 / 29</td>
<td>Kidney deficiency and liver depression</td>
<td>Chinese medicine</td>
<td>TCMS; pregnancy rate, FSH, LH, FSH/LH, AMH, OV, AFC, AE</td>
</tr>
<tr>
<td>Liu LQ et al. 2016</td>
<td>Female</td>
<td>30 / 30</td>
<td>35 / 35</td>
<td>Artificial menstrual cycle (estradiol valerate / medroxyprogesterone)</td>
<td>Chinese medicine</td>
<td>TCMS; pregnancy rate, FSH, LH, FSH/LH, AMH, OV, AFC, AE</td>
</tr>
<tr>
<td>Tang S et al. 2017</td>
<td>Female</td>
<td>20±20</td>
<td>30 / 30</td>
<td>Kidney deficiency and liver depression</td>
<td>Chinese medicine</td>
<td>TCMS; pregnancy rate, FSH, LH, FSH/LH, AMH, OV, AFC, AE</td>
</tr>
<tr>
<td>Zhang JN 2010</td>
<td>Female</td>
<td>30/32</td>
<td>30±30</td>
<td>Kidney deficiency and liver depression</td>
<td>Chinese medicine</td>
<td>TCMS; pregnancy rate, FSH, LH, FSH/LH, AMH, OV, AFC, AE</td>
</tr>
<tr>
<td>Zheng WP et al. 2015</td>
<td>Female</td>
<td>30±30</td>
<td>20 / 20</td>
<td>Kidney deficiency and liver depression</td>
<td>Chinese medicine</td>
<td>TCMS; pregnancy rate, FSH, LH, FSH/LH, AMH, OV, AFC, AE</td>
</tr>
<tr>
<td>Zhou YC et al. 2016</td>
<td>Female</td>
<td>30/30</td>
<td>30±30</td>
<td>Kidney deficiency and liver depression</td>
<td>Chinese medicine</td>
<td>TCMS; pregnancy rate, FSH, LH, FSH/LH, AMH, OV, AFC, AE</td>
</tr>
</tbody>
</table>

Notes: AE: adverse events; AFC: antral follicle count; AMH: anti-Müllerian hormone; E: estradiol; FSH: follicle-stimulating hormone; KI: Kupperman index; LH: luteinizing hormone; OV: ovarian volume; RI: resistance index; TCMSS: Traditional Chinese Medicine symptom score.

**Methods**

**AFC**

AFC was examined in five studies and three meta-analyses could be combined for meta-analysis. The studies showed heterogeneity ($I^2 = 0.03 < 0.1$), so the random-effect model was applied. The results indicated that the AFC of the TCM group was significantly higher than that of the WM group ($MD = 0.55$, $95\% CI = 0.05$ to $1.04$, $155$ participants, three trials) (Figure 5).

One study divided participants into subgroups according to whether the AFC was normal ($\geq 4$) or not ($< 4$) before treatment, and observed changes in AFC in the different subgroups. The criterion used was different from ours, so we did not analyze this index here.

Another study counted the AFC of each side ovary respectively. The baselines of both sides in TCM group and WM group were consistent before treatment. After treatment, the left ovary AFC was $5.33 \pm 1.49$ in the experimental group ($n = 30$) and $3.20 \pm 1.54$ in the control group ($n = 30$), and this difference was significant ($P < 0.0001$). The right ovary AFC after treatment was $5.76 \pm 1.02$ in the experimental group and $5.15 \pm 1.23$ in the control group, and this difference was also significant ($P = 0.04 < 0.05$).
Liu LQ et al. / Systematic Review

Effects of interventions: TCMSS

The TCMSS of the kidney deficiency and liver depression pattern was examined in five studies.\(^\text{23,27-30}\) Of these, one study\(^\text{28}\) reported this incompletely so it was not included in the Meta-analysis, another study\(^\text{27}\) examined TCM combined with WM versus WM, and three studies\(^\text{23,29,30}\) examined TCM versus WM. There was no heterogeneity in these three studies \( (P = 0.62 > 0.1, I^2 = 0\% < 50\%) \), so the fixed-effect model was applied. The results indicated that the TCMSS was significantly lower in the TCM group than in the WM group \( (MD = -2.39, 95\% CI = -3.83 \text{ to } -0.94, 160 \text{ participants, three trials}) \), indicating that TCM was much more effective than WM in reducing TCMSS for DOR patients with kidney deficiency and liver depression pattern \( (P = 0.001 < 0.01) \). The results of the study that assessed the TCMSS in a TCM combined with WM group versus a WM group showed that the TCMSS in an integrated therapy group was much lower \( (MD = -2.93, 95\% CI = -4.89 \text{ to } -0.97, 60 \text{ participants, one trial}) \), indicating that integrated therapy was more effective than WM \( (P = 0.003 < 0.01) \). A meta-analysis of these four studies showed that the TCMSS was much lower in the experimental group than in the control group \( (MD = -2.58, 95\% CI = -3.74 \text{ to } -1.42, 220 \text{ participants, four trials}) \), indicating the effectiveness of TCM in this area \( (P < 0.0001) \) (Figure 6).

Effects of interventions: effective rate of TCMSS

The effective rate of the TCMSS for the kidney deficiency and liver depression pattern was examined in five studies.\(^\text{23,27-30}\) One study\(^\text{28}\) was not included in the Meta-analysis because of its TCMSS data was incomplete. In the rest four studies, one\(^\text{27}\) examined TCM combined with WM versus WM, and the other three\(^\text{23,29,30}\) examined TCM versus WM. These three studies showed no heterogeneity \( (P = 0.42 > 0.1, I^2 = 0\% < 50\%) \), so the fixed-effect model was used. The results showed that the effective rate of TCMSS in the TCM group was significantly higher than that in the WM group \( (RR = 1.18, 95\% CI = 1.02 \text{ to } 1.36, 160 \text{ participants, three trials}) \), indicating that TCM was more effective than WM for DOR patients with kidney deficiency and liver depression pattern \( (P = 0.02 < 0.05) \). The results of the study that examined the effective
rate of TCMSS in a TCM combined with WM group and a WM group showed no difference in the effective rate between the two groups ($RR = 1.13$, $95\% CI: 0.89$ to $1.44$, $60$ participants, one trial, $P = 0.32 > 0.05$). A meta-analysis of these four studies showed a higher effective rate of TCMSS for the experimental group than for the control group ($RR = 1.17$, $95\% CI: 1.03$ to $1.32$, $220$ participants, four trials), indicating the superiority of TCM in this area ($P = 0.01$) (Figure 7).

Effects of interventions: FSH/LH

The FSH/LH ratio was examined in four studies. As one study divided participants into subgroups according to whether FSH/LH ratios were normal ($\leq 3.6$) or not ($> 3.6$) before treatment and observed ratio changes in different subgroups, the Meta-analysis was conducted using data from the other three studies. Of these three, one examined TCM combined with WM versus WM, and the other examined TCM versus WM. These latter two studies showed significant heterogeneity ($P = 0.08 < 0.1$, $I^2 = 67\% > 50\%$, so we used a random-effect model. When the two studies examining TCM versus WM were combined, the results showed no significant difference between the two groups in FSH/LH ratio ($MD = 0.01$, $95\% CI: -0.95$ to $0.96$, $115$ participants, two trials, $P = 0.99 > 0.05$).

The results of the study examining TCM combined with WM versus WM showed that the FSH/LH ratio in the integrated therapy group was a little higher than that in the WM group ($MD = 0.27$, $95\% CI: 0.01$ to $0.53$, $60$ participants, one trial), suggesting a higher effectiveness in the control group ($P = 0.04 < 0.05$). A meta-analysis of these four studies found no significant difference in the FSH/LH ratio between the experimental group and the control group ($MD = 0.1$, $95\% CI: -0.42$ to $0.62$, $175$ participants, three trials) (Figure 8).

We did not analyze the index in the remaining study that was not included in the Meta-analysis, as it used $3.6$ as the cutoff for differentiating normal or abnormal FSH/LH ratios, which was inconsistent with the criterion we set.

Effects of interventions: FSH at 3 months post-treatment

Two studies examined FSH levels at 3 months post-treatment. These showed little heterogeneity ($P = 0.16 > 0.1$, $I^2 = 0.48\% < 50\%$), so the fixed-effect model was used. The results indicated that FSH at 3 months post-treatment was significantly lower in the TCM group than in the WM group ($MD = -4.77$, $95\% CI: -8.00$ to $-1.55$, $297$ participants, three trials) (Figure 9).
CI – 6.69 to –3.45, 137 participants, two trials), indicating a superior effect of TCM in this area (P < 0.0001) (Figure 9).

Adverse events
Six studies recorded adverse events.23,24,27–30 Most adverse events occurred in the control groups and involved nausea27,28 (n = 6) and stomachache27,28 (n = 2). One study23 included brief reports of the following adverse events (n = 3): dizziness, nausea, loss of appetite, and distending breast pain, all of which occurred in the control group. Two studies reported adverse events in the experimental group: nausea27 (n = 1) and diarrhea28 (n = 1). The other three studies reported no adverse events.

Publications bias assessment
Funnel plots were constructed to evaluate the bias in the included studies (Figure 10). The FSH funnel graph showed no obvious bias.

Sensitivity analysis
A sensitivity analysis was conducted on FSH, FSH/LH, AFC, TCMSS, and effective rate of TCMSS, which were reported in more than two studies. Studies were divided into different subgroups according to the type of literature (journal articles or academic dissertations) and the combined effect size in different subgroups was compared. There were no significant subgroup differences in FSH and TCMSS. There were subgroup differences in combined effect size for FSH/LH.
ratios, AFC, and the effective rate of TCMSS (Figures 11-13).

**DISCUSSION**

We searched seven databases and found nine RCTs examining the effectiveness of kidney-tonifying and liver-regulating therapy on DOR. The results show that the effectiveness of kidney-tonifying and liver-regulating therapy is not inferior to that of hormone therapy for DOR.

Most of the studies indicated that the effectiveness of TCM is comparable to that of WM in improving sexual hormone levels (FSH, AMH, and FSH/LH). A combination of TCM and WM\(^2\) seemed much more effective in reducing FSH than only WM. TCM was more effective in improving AFC levels. This may indicate that TCM can fundamentally improve or stimulate ovarian reserve. These outcomes are important indicators when diagnosing DOR, so their improvement suggests that TCM has some effectiveness for DOR.

TCM showed some superiority over WM in improving TCMSS and the effective rate of TCMSS. The symptom score scales are based on the clinical features of the kidney deficiency and liver depression pattern. Tonifying-kidney and regulating-liver therapy may be more effective than WM in relieving clinical symptoms in patients with kidney deficiency and liver depression, which indicates the superiority of syndrome differentiation and treatment. We noticed that TCM also has an advantage in terms of long-term effectiveness (as in FSH tested at 3 months post-treatment). However, perhaps owing to clinical difficulties, there are not many studies in this area. We found only two studies that examined this outcome, which affects the validity of the present findings. Future clinical or experimental studies are needed to focus on long-term outcomes.

The assessment of the adverse effect reporting of the included studies suggests that TCM is safer than WM. The funnel plots showed no obvious bias. The sensitivity analysis indicated instability in the FSH/LH, AFC, and effective rate of TCMSS findings, which may indicate that TCM may have some advantages in reducing FSH levels compared to WM. However, further research is needed to confirm these findings.
result from the difference in literature sources. In addition, there was a publication bias. The number of studies included in the Meta-analysis for each outcome ranged from two to eight (i.e., less than ten). The total sample sizes in the meta-analysis for each outcome were insufficient. Additionally, all studies were conducted in hospitals of TCM or integrated medicine on the Chinese mainland. All journal articles were written in Chinese and published in Chinese journals. Therefore, the findings showed some geographical bias, language bias, and publishing bias.

None of the included studies used blinding and the allocation concealment was unclear. The sample size of each group in all the studies was less than 35. The quality of included studies was generally low. Although objective indicators have been selected to measure effectiveness, as DOR is manifested as an endocrine disorder that is susceptible to mental factors, blinding is necessary for the objectivity of results. Therefore, more randomized clinical trials with large sample sizes, multicenter samples, and double-blinding are needed to confirm that TCM tonifying-kidney and regulating-liver therapy is more effective than WM in DOR treatment.

In conclusion, the effectiveness of tonifying-kidney and regulating-liver therapy is more effective than WM in DOR treatment. Further well-designed RCTs are required to confirm these findings.

REFERENCES


