Menopause: The Journal of The North American Menopause Society
Vol. 26, No. 1, pp. 78-93
DOI: 10.1097/GME.0000000000001161

OPEN

Serum androgen profiles in women with premature ovarian insufficiency: a systematic review and meta-analysis

Midhun Soman, MS,[1] Li-Cong Huang, MD,[1] Wen-Hui Cai, MS,[1] Jun-Bi Xu, MS,[1] Jun-Yao Chen, MD,[1] Ren-Ke He, MS,[1] Heng-Chao Ruan, MS,[1,2,3] Xiang-Rong Xu, MD,[1,2,3] Zhi-Da Qian, MD,[1,2,3,4] and Xiao-Ming Zhu, MD, PhD[1,2,3]

Abstract

Objective: This meta-analysis aims to investigate serum androgen profiles (testosterone, dehydroepiandrosterone sulfate, androstenedione, and sex hormone-binding globulin) in women with premature ovarian failure and to establish if there is evidence of diminished androgen levels in these women.

Methods: Various Internet sources of PubMed, Cochrane library, and Medline were searched systematically until February, 2018. Out of a pool of 2,461 studies, after applying the inclusion/exclusion criterion, 14, 8, 10, and 9 studies were chosen for testosterone, dehydroepiandrosterone sulfate, androstenedione, and sex hormone-binding globulin, respectively, for this meta-analysis. The effect measure was the standardized mean difference with 95% confidence interval (95% CI) in a random-effects model.

Results: The testosterone concentrations in premature ovarian insufficiency were compared with fertile controls: stamdard mean difference (IV, random, 95% CI) −0.73 [−0.99, −0.46], P value < 0.05. The dehydroepiandrosterone sulfate concentrations in premature ovarian insufficiency compared to fertile controls: standard mean difference (IV, random, 95% CI) −0.65 [−0.92, −0.37], P value < 0.05. Androstenedione in premature ovarian insufficiency were compared with fertile controls: standard mean difference (IV, random, 95% CI) −1.09 [−1.71, −0.48], P value < 0.05. Sex hormone-binding globulin levels did not show statistical significance. The dehydroepiandrosterone sulfate levels were reduced in premature ovarian insufficiency cases, but still showed a higher level than in postmenopausal women.

Conclusions: Women with premature ovarian insufficiency are at risk for decreased concentrations of testosterone, dehydroepiandrosterone sulfate, and androstenedione. Dehydroepiandrosterone sulfate levels were more reduced in postmenopausal controls when compared with premature ovarian insufficiency cases.

Key Words: Androgen – Androstenedione – DHEA-S – Premature ovarian insufficiency – SHBG – Testosterone.

Premature ovarian insufficiency (POI) is defined as a decrease in ovarian function occurring before the age of 40.[1-3] It is characterized by hypergonadotropica-menorrhea with estrogen deficiency.

Premature ovarian insufficiency is occasionally referred to as primary ovarian failure (POF), whereas in some situations, the conditions are not the same.[1-6] Women who suffer from POI can have occasionally irregular menstrual cycles or incidental periods for a considerable length of time and may even get to conceive a child.[1-6] Whereas women suffering from POF usually stop having periods and can never be pregnant, POF is generally considered as the end stage of POI.[1-6]

Women with POI/POF not only have a deficiency of estrogens, but they may also have loss of ovarian androgens because of the atrophy of the ovarian cortex.[6-9] Androgens are thought to be one of the fundamental prerequisites for a healthy woman.[7-14] A lack of androgens may lead to symptoms of sexual dysfunction, such as decreased libido, loss of sexual responsiveness, or decreased sexual arousal.[7-14] Other clinical manifestations are a diminished sense of well-being, dysphoric mood, cognitive dysfunction and persistent, and unexplained loss of energy.[7-14]

Received March 29, 2018; revised and accepted May 18, 2018.

From the [1]Department of Obstetrics & Gynecology, Women's Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, P.R. China; [2]The Key Laboratory of Reproductive Genetics (Zhejiang University), Ministry of Education, Hangzhou, P.R. China; [3]The Key Laboratory of Women's Reproductive Health Research of Zhejiang Province, Hangzhou, Zhejiang, P.R. China; and [4]Department of Family Planning, Women's Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, P.R. China.

MS and L-CH contributed equally to this work.

Funding/support: None reported.

Financial disclosure/conflicts of interest: None reported.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website (www.menopause.org).

Address correspondence to: Xiao-Ming Zhu, MD, PhD, and Zhi-Da Qian MD, Department of Obstetrics & Gynecology, Women's hospital, School of Medicine, Zhejiang University, 1 Xueshi Rd. Hangzhou, 310006 Zhejiang, P.R. China.
E-mail: ntzhuxm@zju.edu.cn, qianada@zju.edu.cn

The role of androgens in maintaining a woman's health has been receiving increasing attention, but still, there is considerable controversy relating to the role of androgens in women, and it's clinical significance.[4,13-21] Be as it may, there are guidelines regarding the diagnosis of androgen deficiency and the relative androgen replacement therapy in women.[4,13-21] POI/POF maybe highly suspected as a cause of clinically significant androgen deficiency.[5,22-24]

It has not been clear if there is a decrease in androgen levels for women with POI/POF. It is clear that, at present, the state of spontaneous POI/POF is different from the natural menopausal process.[1,21-25] Many studies have shown evidence that there is a gradual decline in circulating serum androgen levels as women age and that there is no independent effect of natural menopause causing a further reduction.[19,20,21,26-28]

If there is indeed a clinically significant reduction in androgen levels, androgen replacement therapy could be considered to augment the standard estrogen-based hormone therapy for women with POI/POF, but this needs to be further investigated.[1]

Some studies have shown that there was a positive correlation between serum androgen concentration in women with POI/POF to age. POI/POF is a pathologic condition in which women have lower serum estradiol levels as compared with other women of similar age.[1] It is yet to be known if women with POI/POF present with lower androgen levels when compared with menopausal women. Also, there is uncertainty as to whether POI/POF is a cause of clinically significant androgen-deficiency.[5,22-24] There have been many individual studies comparing the androgen levels in women with POI/POF. However, the results vary for each study due to study methods employed and the use of different assays for measurement.

This systematic review and meta-analysis aims to incorporate the data from all these studies collected using different methodologies by standardizing the unit of measurement and making use of a standardized mean difference (SMD) to calculate the overall effect, and to verify if there is indeed a significant reduction in androgen concentrations in women with POI/POF compared with a healthy control group.

METHODS

Eligibility criteria

In this meta-analysis, all published studies in which serum androgen concentrations including total testosterone (T), androstenedione (A), dehydroepiandrosterone sulfate (DHEA-S), sex hormone-binding globulin (SHBG) were described for women with POI/POF and compared with healthy controls, were considered eligible. The predefined criteria for inclusion were as follows: POI/POF had to be defined as the disappearance of periods (amenorrhea) for the past 3 months; the disappearance of periods must have occurred even before the age of 40 is attained; in addition, the participants of the study must have had elevated follicle-stimulating hormone (FSH) levels and reduced estradiol concentrations; controls were required to be women without POI/POF. The predefined exclusion criteria were as follows: the studies were excluded if the participants had hyperandrogenemia, bilateral salpingo-oophorectomy (BSO), or gonadotoxic treatment performed after menopause; if the participants used hormone therapy, the study was excluded; studies including participants other than females were excluded; studies, which focused on chromosomally abnormal POI/POF women, were excluded; reviews, case reports, letters to the editor, and conference papers were not considered for inclusion. Only English language articles were examined. Study methodology was not considered as an exclusion criterion.

The study selection processes were carried out in two phases. First and foremost, screening of titles and abstract was done to meet the inclusion criterion by two separate researchers. The consensus among themselves resolved any disagreements.

Search strategy

The searched databases were PubMed, Cochrane library, and Medline. The time range was inception to February, 2018. Primarily the studies were evaluated by reading the titles and abstracts. The search terms were menopause, premature, ovarian, insufficiency, failure, POI, POF, testosterone, total testosterone, DHEA-S, androstenedione, and SHBG. The search syntax for PubMed central was:(''menopause, premature''[MeSH Terms] OR (''menopause''[All Fields] AND ''premature''[All Fields]) OR ''premature menopause''[All Fields] OR (''premature''[All Fields] AND ''ovarian''[All Fields] AND ''failure''[All Fields]) OR ''premature ovarian failure''[All Fields]) AND (''androgens''[All Fields] OR ''androgens''[MeSH Terms] OR ''androgens''[All Fields] OR ''androgen''[All Fields]). The search was modified for other databases used. Reference lists of relevant review articles were searched for potentially eligible studies. When required, authors were contacted for additional information.

Study selection

Studies were scrutinized by two reviewers independently according to the predefined inclusion-exclusion criteria. The final decisions about the included articles were made by examining the full articles.

Quality assessment and data extraction

Two independent reviewers assessed the risk of bias with the Newcastle-Ottawa Scale for meta-analysis of observational studies.[29] Newcastle-Ottawa Scale for meta-analysis of observational studies grades studies based on three factors: selection, comparability, and exposure. The criteria for these assessments were as follows: is the case definition adequate? representativeness of the cases; selection of controls; definition of controls; comparability of cases and controls based on the design or analysis; ascertainment of exposure; the same method of ascertainment for cases and controls; and nonresponse rate. Two reviewers extracted data from all eligible articles independently. Data that were collected included the year of publication, study design, country, age, body mass index (BMI), number of participants, mean and standard deviation (SD) of testosterone concentrations (ng/dL), androstenedione (ng/dL), DHEA-S (μg/mL), and SHBG (nmol/L) for both cases and controls. Selection can have a maximum of

4*s, comparability a maximum of 2*s based on the two important confounding factors. In our study, we chose age and BMI as the two factors. Exposure can have a maximum of 3*s. So, altogether, 9*s are the highest that can be obtained. The high and low score ranges are arbitrarily given. Here the lowest score was 2*s and highest was 8*s. So we used a scale of 0 to 2*s being poor quality, 3 to 5*s as fair quality, and 6 to 9*s being higher quality. A study is considered good if it can score at least 1* in each category, and more *s means better quality generally.

Statistical analysis

In this meta-analysis, the SMD with 95% confidence intervals (CIs) were solved to determine T concentrations, A concentrations, DHEA-S, and SHBG between POI/POF cases and controls. Random-effect model (REM)[30] was used to calculate the pooled effect size with 95% CI. Analyses were conducted using Revman 5.3.[31] Heterogeneity among studies was assessed using chi-square and quantified with the I^2 index. If the I^2 index is 0% to 40%, it means low heterogeneity and might not be important; if the I^2 index is 30% to 60%, it means moderate heterogeneity; if the I^2 index is 50% to 90%, it means substantial heterogeneity; and if the I^2 index is 75% to 100%, it means considerable heterogeneity.[32] Funnel plot and Egger's test[33] were used to evaluate the existence of publication bias. A subgroup analysis was done to investigate the source of heterogeneity between studies, and sensitivity analysis with leave-one-out method was performed to assess robustness of the meta-analysis. The leave-one-out method was achieved by removing one study at a time and measuring the pooled estimate.

Main study characteristics

Various system searches resulted in 2,464 studies discussing POI/POF. The duplicate of the studies was eliminated by use of reference manager; this yielded 1,621 studies. By use of selection criteria, abstract, and screening of the title generated in the differentiation of 192 studies, text papers were retrieved from this search and were analyzed by the selection criteria. During the full-text review, 178 articles were removed. Out of them, 98 studies were removed from the list of studies because of not fulfilling the inclusion criteria. An additional 80 studies were removed from the list of studies because POI/POF diagnosis failed to meet prescribed standards. Finally, 14 studies were included for this meta-analysis (Fig. 1). Out of these 14 studies, 14, 8, 10, and 9 studies were included for total T, DHEA-S, A, and SHBG, respectively. Study characteristics for each androgen are mentioned in Tables 1-4.

Measurement of risk bias

The risk of bias assessment for studies selected for the comparison of androgen concentrations in POI/POF versus

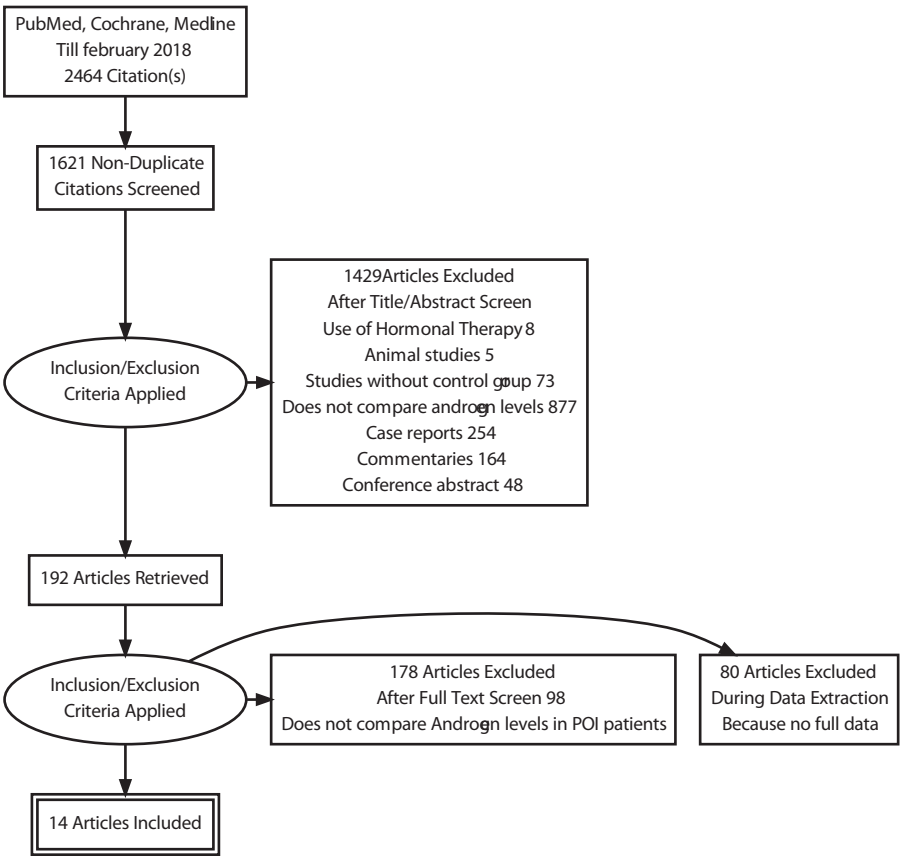

FIG. 1. Search strategy, PRISMA flow chart for meta-analysis on total androgen concentrations.

TABLE 1. *Study characteristics for meta-analysis on testosterone levels*

					POF				Control				Postmenopausal women			
No.	Study	Year	Country	Study design	Age mean ± SD (y)	BMI mean ± SD (kg/m^2)	n	T mean ± SD (ng/dL)	Age mean ± SD (y)	BMI Mean ± SD (kg/m^2)	n	T mean ± SD (ng/dL)	Age mean ± SD (y)	BMI mean ± SD (kg/m^2)	n	T mean ± SD (ng/dL)
1	Bermudez et al[25]	1993	Mexico	Case-control, cross-sectional	Range: 20 to 34	Range: 20 to 25	7	14.41 ± 8.93	Range: 27 to 29	Range: 20 to 25	6	30.84 ± 19.88		NA		
2	Hartmann et al[24]	1997	Austria	Case-control, cross-sectional	28.7 ± 4.9	22.5 ± 3.6	33	25.00 ± 16.0	28.3 ± 4.9	21.6 ± 2.6	33	43.00 ± 24.00	53.1 ± 2.6	24.5 ± 2.6	32	27 ± 13
3	Elias et al[7]	1997	USA	Case-control, cross-sectional	34 ± 4	NA	29	27.67 ± 10.9	34 ± 4	NA	29	27.95 ± 5.48		NA		
4	Doldi et al[37]	1998	Italy	Case-control, cross-sectional	30.2	22.4	25	39.58 ± 11.5	29.4	21.1	18	60.52 ± 17.29		NA		
5	Falsetti et al[38]	1999	Italy	Case-control, cross-sectional	32.6 ± 7.3	22.9 ± 3.8	40	28.82 ± 8.65	35 ± 3.5	22.2 ± 2.2	30	43.23 ± 11.53		NA		
6	Benetti-Pinto et al[21]	2005	Brazil	Case-control, cross-sectional	34.4 ± 5.2	24.7 ± 5	30	22.8 ± 10.2	34.5 ± 5.5	24.4 ± 4.6	30	26.6 ± 12.6	55.1 ± 3.9	NA	30	17.5 ± 7.7
7	Kalantaridou et al[34]	2006	USA	Case-control, cross-sectional	32.1 ± 5.5	23.2 ± 3.1	130	23.4 ± 9.3	30.3 ± 7.1	23 ± 2.7	65	30.9 ± 10.5		NA		
8	van der Stege et al[35]	2008	Netherlands	Case-control, cross-sectional	35.8 ± 4.9	23.5 ± 3.4	27	64.43 ± 22.7	35 ± 4.7	24 ± 4.6	63	85.59 ± 22.48		NA		
9	Janse et al[41]	2011	Netherlands	Case-control, cross-sectional	37.6 ± 11.77	23.4 ± 5.05	208	32.00 ± 12.0	33.1 ± 4.3	23.4 ± 4.9	45	39.00 ± 17.00		NA		
10	Gulhan et al[36]	2012	Turkey	Case-control	36.8 ± 1.8	25.9 ± 4.9	47	31.5 ± 12.5	36 ± 2.3	24.8 ± 4.6	60	34.2 ± 14.5		NA		
11	Ates et al[40]	2014	Turkey	Case-control	35.23 ± 4.58	25.79 ± 4.1	56	46 ± 121	35.4 ± 4.49	26.0 ± 3.80	59	49.00 ± 90.00		NA		
12	Daan et al[43]	2011	Netherlands	Case-control, cross-sectional	35.00 ± 1.9	23.0 ± 1.3	170	20 ± 3	33.0 ± 1.75	23.4 ± 1.05	170	23 ± 3	55.8 ± 1.2	26.3 ± 1.2	170	23.05 ± 5.76
13	Szlendak-Sauer et al[39]	2016	Poland	Case-control	30.5 ± 6.3	NA	98	39.00 ± 24.0	29.4 ± 4	23.2 ± 2.5	75	45.00 ± 19.00		NA		
14	Florence et al[42]	2016	Kuwait	Case-control	28.8 ± 6.4	27.8 ± 4.8	31	35.00 ± 12.0	29.2 ± 6.8	28.2 ± 4.4	31	63.00 ± 17.00		NA		

BMI, body mass index; NA, not available; POF, premature ovarian failure; T, testosterone concentration.

TABLE 2. *Study characteristics for meta-analysis on DHEA-S levels*

					POF				Control				Postmenopausal women			
No.	Study	Year	Country	Study design	Age mean ± SD (y)	BMI mean ± SD (kg/m^2)	n	DHEA-S mean ± SD (mcg/mL)	Age mean ± SD (y)	BMI mean ± SD (kg/m^2)	n	DHEA-S[d] mean ± SD (mcg/mL)	Age mean ± SD (y)	BMI mean ± SD (kg/m^2)	n	DHEA-S mean ± SD (mcg/mL)
1	Hartmann et al[24]	1997	Austria	Case-control, cross-sectional	28.7 ± 4.9	22.5 ± 3.6	33	1.75 ± 0.99	28.3 ± 4.9	21.6 ± 2.6	33	2.15 ± 1.19	53.1 ± 2.6	24.5 ± 2.6	32	1.43 ± 0.87
2	Elias et al[7]	1997	USA	Case-control, cross-sectional	34 ± 4	NA	29	1.25 ± 0.76	34 ± 4	NA	29	1.66 ± 0.82		NA		
3	Doldi et al[37]	1998	Italy	Case-control, cross-sectional	30.2	22.4	25	0.56 ± 0.53	29.4	21.1	18	1.20 ± 1.06		NA		
4	Falsetti et al[38]	1999	Italy	Case-control, cross-sectional	32.6 ± 7.3	22.9 ± 3.8	40	1.1 ± 0.6	35 ± 3.5	22.2 ± 2.2	30	1.9 ± 0.6		NA		
5	Benetti-Pinto et al[21]	2005	Brazil	Case-control, cross-sectional	34.4 ± 5.2	24.7 ± 5	30	0.72 ± 0.33	34.5 ± 5.5	24.4 ± 4.6	30	0.83 ± 0.42	55.1 ± 3.9	NA	30	0.43 ± 0.2
6	van der Stege et al[35]	2008	Netherlands	Case-control, cross-sectional	35.8 ± 4.9	23.5 ± 3.4	27	1.8 ± 0.81	35 ± 4.7	24 ± 4.6	63	2.22 ± 1.22		NA		
7	Daan et al[43]	2011	Netherlands	Case-control, cross-sectional	35.00 ± 1.9	23.0 ± 1.3	170	1.28 ± 3	33.0 ± 1.75	23.4 ± 1.05	170	1.47 ± 0.16	55.8 ± 1.2	26.3 ± 1.2	170	0.97 ± 0.17
8	Szlendak-Sauer et al[39]	2016	Poland	Case-control	30.5 ± 6.3	NA	98	2.17 ± 1.14	29.4 ± 4	23.2 ± 2.5	75	2.65 ± 1.06		NA		

BMI, body mass index; DHEA-S, dehydroepiandrosteronesulfate; NA, not available; POF, premature ovarian failure.

TABLE 3. *Sudy characteristics for meta-analysis on androstenedione levels*

					POF				Control				Postmenopausal women			
No.	Study	Year	Country	Study design	Age mean ± SD (y)	BMI mean ± SD (kg/m^2)	n	A mean ± SD (ng/dL)	Age mean ± SD (y)	BMI mean ± SD (kg/m^2)	n	A mean ± SD (ng/dL)	Age mean ± SD (y)	BMI mean ± SD (kg/m^2)	n	A mean ± SD (ng/dL)
1	Hartmann et al[24]	1997	Austria	Case-control, cross-sectional	28.7 ± 4.9	22.5 ± 3.6	33	157.00 ± 66.00	28.3 ± 4.9	21.6 ± 2.6	33	215 ± 93.00	53.1 ± 2.6	24.5 ± 2.6	32	138 ± 47
2	Elias et al[7]	1997	USA	Case-control, cross-sectional	34 ± 4	NA	29	88.13 ± 32.14	34.00 ± 4.00	NA	29	119.36 ± 39.55		NA		
3	Doldi et al[37]	1998	Italy	Case-control, cross-sectional	30.2	22.4	25	86.00 ± 42.4	29.4	21.1	18	150.00 ± 84.8		NA		
4	Falsetti et al[38]	1999	Italy	Case-control, cross-sectional	32.6 ± 7.3	22.9 ± 3.8	40	100.00 ± 20.00	35.00 ± 3.5	22.2 ± 2.2	30	190.00 ± 50.00		NA		
5	Benetti-Pinto et al[21]	2005	Brazil	Case-control, cross-sectional	34.4 ± 5.2	24.7 ± 5	30	130.00 ± 50.00	34.5 ± 5.5	24.4 ± 4.6	30	160.00 ± 80.00	55.1 ± 3.9	NA	30	100 ± 40
6	van der Stege et al[35]	2008	Netherlands	Case-control, cross-sectional	35.8 ± 4.9	23.5 ± 3.4	27	107.4 ± 42.38	35 ± 4.7	24 ± 4.6	63	196.47 ± 84.7		NA		
7	Janse et al[41]	2011	Netherlands	Case-control, cross-sectional	37.6 ± 11.77	23.4 ± 5.05	208	63.86 ± 58.13	33.1 ± 4.3	23.4 ± 4.9	45	94.00 ± 68.66		NA		
8	Daan et al[43]	2011	Netherlands	case-control, cross-sectional	35.00 ± 1.9	23.0 ± 1.3	170	68.73 ± 9.16	33.0 ± 1.75	23.4 ± 1.05	170	100.24 ± 12.88	55.8 ± 1.2	26.3 ± 1.2	170	68.73 ± 10.7
9	Szlendak-Sauer et al[39]	2016	Poland	case-control	30.5 ± 6.3	NA	98	238.00 ± 111.00	29.4 ± 4	23.2 ± 2.5	75	328.00 ± 120.00		NA		
10	Florence et al[42]	2016	Kuwait	Case-control	28.8 ± 6.4	27.8 ± 4.8	31	138.00 ± 63.00	29.2 ± 6.8	28.2 ± 4.4	31	150.00 ± 63.00		NA		

A, androstenedione; BMI, body mass index; NA, not available; POF, premature ovarian failure.

TABLE 4. *Study characteristics for meta-analysis on SHBG levels*

					POF				Control			
No.	Study	Year	Country	Study design	Age mean ± SD (y)	BMI mean ± SD (kg/m^2)	n	**SHBG mean ± SD (nmol/L)**	Age mean ± **SD** (y)	BMI mean ± **SD** (kg/m^2)	n	**SHBG mean ± SD (nmol/L)**
1	Hartmann et al[24]	1997	Austria	Case-control, cross-sectional	28.7 ± 4.9	22.5 ± 3.6	33	53.9 ± 28.5	28.3 ± 4.9	21.6 ± 2.6	33	66.7 ± 34.7
2	Falsetti et al[38]	1999	Italy	Case-control, cross-sectional	32.6 ± 7.3	22.9 ± 3.8	40	47.5 ± 5.1	35 ± 3.5	22.2 ± 2.2	30	62.89 ± 31.55
3	Benetti-Pinto et al[21]	2005	Brazil	Case-control, cross-sectional	34.4 ± 5.2	24.7 ± 5	30	64.1 ± 37.9	34.5 ± 5.5	24.4 ± 4.6	30	67 ± 30.2
4	Kalantaridou et al[34]	2006	USA	Case-control, cross-sectional	32.1 ± 5.5	23.2 ± 3.1	130	102.7 ± 45.6	30.3 ± 7.1	23 ± 2.7	65	106.4 ± 53.5
5	Janse et al[41]	2011	Netherlands	Case-control, cross-sectional	37.6 ± 11.77	23.4 ± 5.05	208	55.00 ± 46.00	33.1 ± 4.3	23.4 ± 4.9	45	60.00 ± 46.75
6	Ates et al[40]	2014	Turkey	Case-control	35.23 ± 4.58	25.79 ± 4.10	56	72.58 ± 61.45	35.4 ± 4.49	26.0 ± 3.80	59	62.89 ± 31.55
7	Daan et al[43]	2011	Netherlands	Case-control, cross-sectional	35.00 ± 1.9	23.0 ± 1.3	170	54.00 ± 18.45	33.0 ± 1.75	23.4 ± 1.05	170	48.5 ± 14.2
8	Szlendak-Sauer et al[39]	2016	Poland	Case-control	30.5 ± 6.3	NA	98	50.48 ± 26.6	29.4 ± 4	23.2 ± 2.5	75	59.3 ± 30.9
9	Florence et al[42]	2016	Kuwait	Case-control	28.8 ± 6.4	27.8 ± 4.8	31	216 ± 64	29.2 ± 6.8	28.2 ± 4.4	31	234 ± 68

BMI, body mass index; NA, not available; POF, premature ovarian failure; SHBG: sex hormone-binding globulin.

TABLE 5. *Assessment of the risk of bias of involved studies using the Newcastle-Ottawa Scale for meta-analysis of observational studies*

	Study	Year	Selection	Comparability	Exposure
1	Benetti-Pinto et al[21]	2005	**	**	*
2	Kalantaridou et al[34]	2006	****	***	*
3	van der Stege et al[35]	2008	***	**	**
4	Gulhan et al[36]	2012	**	**	**
5	Elias et al[7]	1997	*	*	—
6	Doldi et al[37]	1998	**	**	**
7	Bermudez et al[25]	1993	**	—	*
8	Hartmann et al[24]	1997	**	**	**
9	Falsetti et al[38]	1999	**	**	*
10	Szlendak-Sauer et al[39]	2016	**	—	**
11	Ates et al[40]	2014	**	**	**
12	Janse et al[41]	2011	***	**	**
13	Florence et al[42]	2016	**	**	*
14	Daan et al[43]	2015	***	—	**

The criteria for Quality Assessment using Newcastle-Ottawa Scale for meta-analysis.
Selection:
1. Is the case definition adequate? (a) yes, with independent validation*; (b) yes, for example, record linkage or based on self reports; (c) no description.
2. Representativeness of cases: (a) consecutive or obviously representative series of cases*; (b) potential for selection biases or not stated.
3. Selection of controls: (a) community controls*; (b) hospital controls; (c) no description.
4. Definition of controls: (a) no history of disease (endpoint)*; (b) no description of source.
Comparability:
1. Comparability of cases and controls on the basis of the design or analysis: (a) study controls for (select most important factor)*; (b) study controls for any additional factor* (this criterion could be modified to indicate specific control for a second important factor).
Exposure:
1. Ascertainment of exposure: (a) secure record (eg, surgical records)*; (b) structured interview where blind to case/control status*; (c) interview not blinded to case/control status; (d) written self-report or medical record only e) no description.
2. Same method of ascertainment for cases and controls: (a) yes*; (b) no.
3. Nonresponse rate: (a) same rate for both groups*; (b) nonrespondents described; (c) rate different and no designation.
Note-1: The high and low score ranges are arbitrarily given.
Note-2: 0 to 2*s being poor quality, 3 to 5*s as fair quality, 6 to 9*s being higher quality.
Note-3: Maximum score that can be obtained is 9*s.

controls is reported in Table 5 [see Table, Supplemental Digital Content 1, http://links.lww.com/MENO/A333]. All the studies had clearly defined selection criteria for the cases, but for the controls, the selection is not described, although all the studies except one had defined the controls. All the studies except three fulfilled the comparability assessment. Most of the studies had fulfilled the exposure assessment. Altogether, seven studies had a score of more than 6*s, namely that of Kalantaridou et al, van der Stege et al, Gulhan et al, Doldi et al, Hartmann et al, Janse et al, and Ates et al. For sensitivity analysis of testosterone concentrations, the studies by Ates et al, Gulhan et al, and Doldi et al were not included because even though they scored 6*s on the scale, the selection methodology for cases and controls were not clearly mentioned. The studies by Kalantaridou et al, van der Stege et al, and Janse et al scored 7*s or more on the scale and were of the highest methodological quality. The study by Hartmann et al, while scoring only 6*s, had mentioned the selection methodology; it was just not clear if the participants were consecutively chosen, but because it mentioned how they were chosen as opposed to no description in the other three studies, it was included. Also the study by Doldi et al was included in the sensitivity analysis for androstenedione and DHEA-S because we wanted to compare a minimum of three studies to do the sensitivity analysis. The study by Doldi et al compared both androstenedione and DHEA-S, and scored more than 6*s on the Newcastle-Ottawa Scale. In the absence of other higher-quality studies for these two androgens (androstenedione, DHEA-S), we included this study in the sensitivity analysis. The study by Benetti-Pinto et al had no description of how cases and controls were selected. The study by Elias et al did not have age or BMI-matched cases and controls, nor did it have description of how cases and controls were selected. The study by Falsetti et al had no description of how cases and controls were selected, nor did it have description if both cases and controls used similar tests to check for exposure.

RESULTS

Testosterone concentrations

Women suffering from POI/POF have lower testosterone in comparison to the controls according to data retrieved from 14 studies (n = 1,656) (Table 1), SMD (IV, random, 95% CI) −0.73 [−0.99, −0.46] with P value < 0.05 (Fig. 2). There was significant heterogeneity observed in the meta-analysis ($I^2 = 82\%$). With heterogeneity: $\text{Tau}^2 = 0.19$; $\chi^2 = 72.18$, df = 13 ($P < 0.00001$); so, random-effects model was used. The funnel plot was evaluated to check for publication bias. A subgroup analysis was done to verify the source of heterogeneity. The sensitively analysis showed a consistency among the studies.

Subgroup analysis for total testosterone concentrations

The subgroups were divided according to the various assays used for measurement. For the analysis of total testosterone concentrations, there were four subgroups identified. The first subgroup consisted of four studies which directly applied

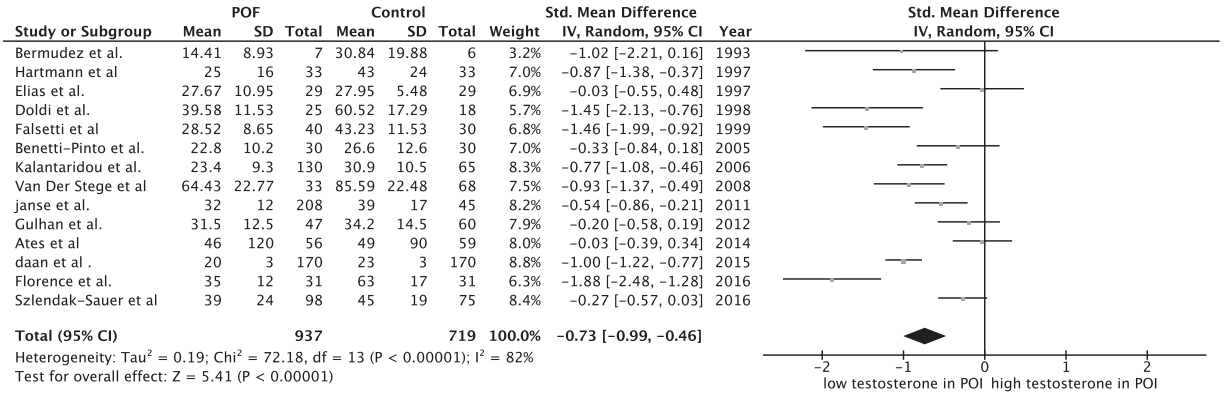

FIG. 2. Forest plot for meta-analysis of 14 comparative studies on total testosterone concentrations in women with POI/POF compared with fertile controls. Squares represent effect of each study, and the diamond represents the overall effect of the study at 95% confidence intervals. The values to the left of the solid line indicate SMD less than 1, decreased concentrations in POI/POF group. The values to the right of the solid line indicate SMD greater than 1, increased concentrations in POF/POI when compared to the control group. POF, primary ovarian failure; POI, premature ovarian insufficiency; SMD, standardized mean difference.

radioimmunoassay (RIA) for the measurement of testosterone concentrations. The second subgroup consisted of five studies which used extraction or chromatography before applying RIA. The third subgroup consisted of four studies which utilized the chemiluminescent immunoassay. And the fourth subgroup consisted of only one study in which it was not clear which assay was used [see Fig., Supplemental Digital Content 2, http://links.lww.com/MENO/A334]

Direct RIA subgroup

The subgroup consisted of four studies (n = 280). This subgroup showed an SMD of −1.13 [−1.44, −0.82] test for overall effect: $Z = 7.23$ ($P < 0.00001$) with substantial decrease in heterogeneity ($\text{Tau}^2 = 0.03$; $\chi^2 = 4.05$, df = 3 [$P = 0.26$]; $I^2 = 26\%$).

Extraction/chromatography RIA subgroup

The second subgroup consisted of studies which used extraction or chromatography before applying RIA. This consisted of five studies (n = 859). This subgroup showed an SMD of −0.66 [−0.99, −0.34], $Z = 4.00$ ($P < 0.0001$). Heterogeneity was still substantial in this subgroup with $\text{Tau}^2 = 0.09$; $\chi^2 = 14.08$, df = 4 ($P = 0.007$); $I^2 = 72\%$. The source of this heterogeneity could be explained by the fact that three of the five studies included in this particular subgroup happened to be lower-quality studies and were not age or BMI-matched. We tried removing these three studies from this subgroup and saw that the heterogeneity had substantially reduced to about 4%.

Chemiluminescent immunoassay subgroup

The subgroup consisted of studies measured using the chemiluminescent immunoassay, and it consisted of four studies (n = 455).This subgroup showed a SMD of −0.20 [−0.38, −0.01], $Z = 2.10$ ($P = 0.04$). This subgroup showed no heterogeneity with $\text{Tau}^2 = 0.00$; $\chi^2 = 1.30$, df = 3 ($P = 0.73$); $I^2 = 0\%$.

Other/nonspecified assay subgroup

The subgroup consisted of only one study (n = 62). This subgroup showed an SMD of −1.88 [−2.48, −1.28], $Z = 6.10$ ($P < 0.00001$).

POI/POF versus fertile controls versus postmenopausal controls

The POI/POF versus fertile controls subgroup showed a SMD (−0.73 [−0.99, −0.46]) with P value <0.05, whereas the POI/POF versus postmenopausal controls (n = 465) showed a SMD of −0.05 [−0.68, 0.58] test for overall effect: $Z = 0.15$ ($P = 0.88$). Heterogeneity: $\text{Tau}^2 = 0.27$; $\chi^2 = 14.92$, df = 2 ($P = 0.0006$); $I^2 = 87\%$. The test for subgroup differences: $\chi^2 = 3.77$, df = 1 ($P = 0.05$), $I^2 = 73.5\%$, indicating substantial heterogeneity between these subgroups (see Fig., Supplemental Digital Content 3, http://links.lww.com/MENO/A335).

Sensitivity analysis for testosterone concentrations

Sensitivity analysis was first conducted using the leave-one-out method by systematically omitting one study at a time, and the results were robust. Then, four studies (n = 615) that scored 6 or greater on the Newcastle-Ottawa Scale, considered as the best-quality studies, were chosen out of the 14 studies. Seven studies had a score of more than 6, namely the studies by Kalantaridou et al, van der Stege et al, Gulhan et al, Doldi et al, Hartmann et al, Janse et al, and Ates et al. The studies by Ates et al, Gulhan et al, and Doldi et al were not included, because even though they scored 6 on the scale, the selection methodology for cases and controls were not clearly mentioned. The studies by Kalantaridou et al, van der Stege et al, and Janse et al scored 7 or more on the scale and were of the highest methodological quality. The study by Hartmann et al, while scoring only 6, mentioned the selection methodology; it was just that it was not clear if the participants were consecutively chosen in this study, but because it mentioned how they were chosen as opposed to no description in the other three studies, it was included. These studies used

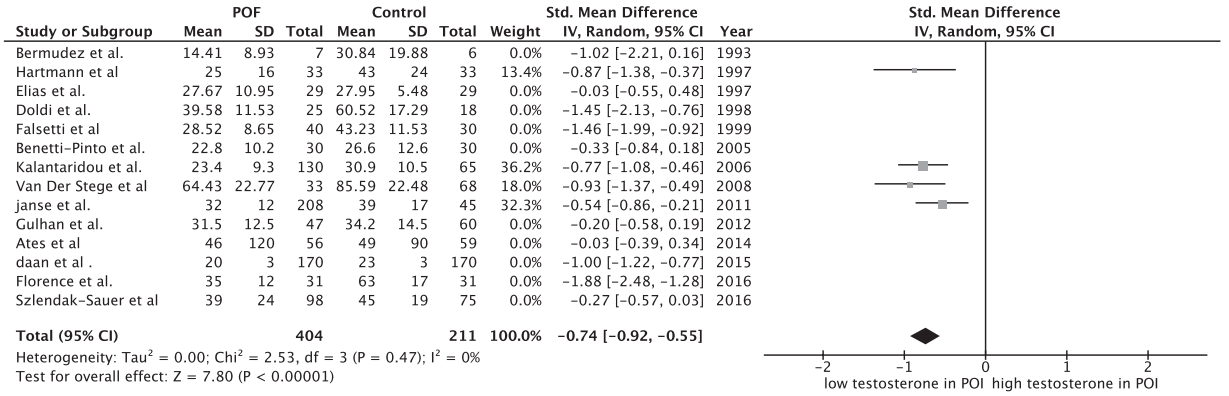

FIG. 3. Sensitivity analysis for total testosterone concentrations in women with POI/POF compared with fertile controls. Four studies (n = 615) that scored >6 on the Newcastle-Ottawa Scale, considered as the best quality studies, were chosen out of the 14 studies. POF, primary ovarian failure; POI, premature ovarian insufficiency.

for the sensitivity analysis were the ones with the best methodological quality and considered having the least amount of bias, and all these studies were age and BMI-matched; this showed an SMD of −0.74 [−0.92, −0.55], $Z = 7.80$ ($P < 0.00001$). With a drop in heterogeneity to 0%: $\text{Tau}^2 = 0.00$; $\chi^2 = 2.53$, df = 3 ($P = 0.47$); $I^2 = 0\%$ (Fig. 3).

Dehydroepiandrosterone sulfate concentrations

Women suffering from POI/POF have lower DHEA-S in comparison with the controls according to data retrieved from eight studies (n = 911) (Table 2) (SMD [IV, random, 95% CI] −0.65 [−0.92, −0.37]), with P value <0.05. There was significant heterogeneity observed in the meta-analysis ($I^2 = 71\%$), with heterogeneity: $\text{Tau}^2 = 0.11$; $\chi^2 = 24.44$, df = 7 ($P = 0.0010$); so, random-effects model was used. The funnel plot was evaluated to check for publication bias. A subgroup analysis was done to verify the source of heterogeneity. The sensitively analysis showed a consistency among the studies (Fig. 4).

Subgroup analysis for DHEA-S concentrations

The subgroups were divided according to the various assays used for measurement. For analysis of DHEA-S concentrations, there were two subgroups identified. The first subgroup consisted of six studies which directly applied RIA for the measurement of DHEA-S concentrations. The second subgroup was the studies in which the assays used were either not mentioned or needed more clarification. This consisted of two studies (see Fig., Supplemental Digital Content 4, http://links.lww.com/MENO/A336).

Direct RIA subgroup

The subgroup consisted of six studies (n = 398). This subgroup showed a SMD of −0.60 [−0.90, −0.29], $Z = 3.79$ ($P = 0.0002$) with a decrease in heterogeneity: $\text{Tau}^2 = 0.08$; $\chi^2 = 10.94$, df = 5 ($P = 0.05$); $I^2 = 54\%$.The source of this heterogeneity could be explained by the fact that three of the six studies included in this particular subgroup happened to be studies with lower methodological quality. The study by Benetti-Pinto et al had no description of how cases and controls were selected. The study by Elias et al did not have age or BMI-matched cases and controls, nor did it have a description of how cases and controls were selected. The study by Falsetti et al had no description of how cases and controls were selected, nor did it have a description if both cases and controls used similar tests to check for

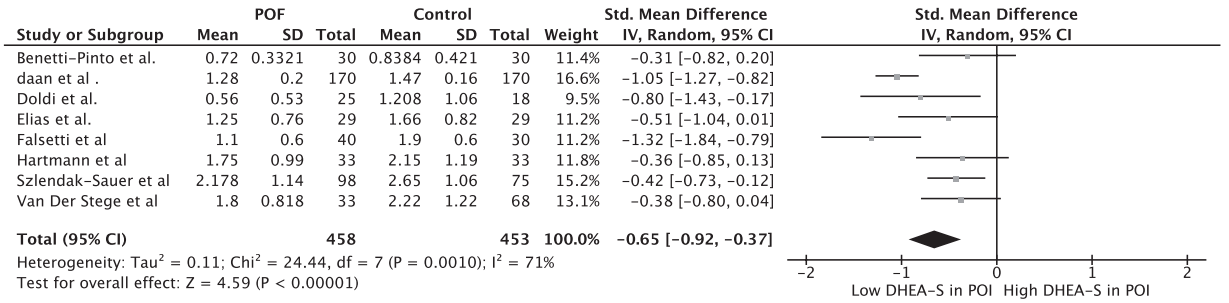

FIG. 4. Forest plot for meta-analysis of eight comparative studies on DHEA-S concentrations in women with POI/POF compared with fertile controls. Squares represent effect of each study, and the diamond represents the overall effect of the study at 95% confidence intervals. The values to the left of the solid line indicate SMD less than 1, decreased concentrations in POI/POF group. The values to the right of the solid line indicate SMD greater than 1, increased concentrations in POF/POI compared to the control group. DHEA-S, dehydroepiandrosterone sulfate; POF, primary ovarian failure; POI, premature ovarian insufficiency; SMD, standardized mean difference.

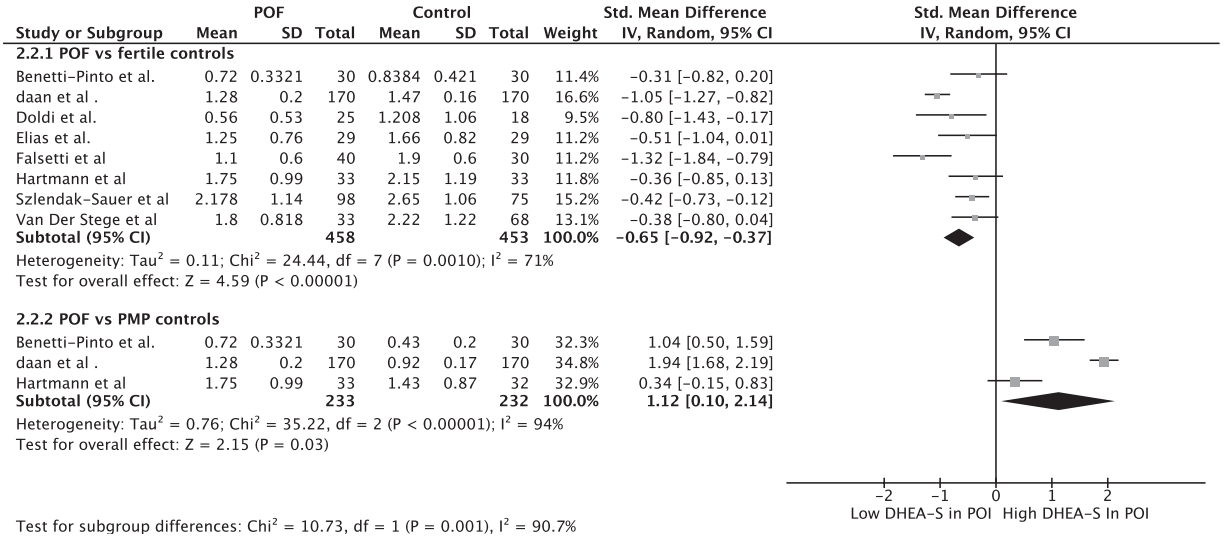

FIG. 5. Subgroup analysis for DHEA-S concentrations in women with POI/POF compared with postmenopausal controls (n = 465), indicating that the DHEA-S levels are much reduced in POI/POF controls, but still show a higher concentration than DHEA-S in postmenopausal women. DHEA-S, dehydroepiandrosterone sulfate; POF, primary ovarian failure; POI, premature ovarian insufficiency.

exposure. We tried removing these studies from this subgroup and saw that the heterogeneity had substantially reduced to about 0%.

Other/nonspecified assay subgroup

The subgroup consisted of only two studies (n = 513). This subgroup showed a SMD of −0.74 [−1.35, −0.14], $Z = 2.39$ ($P = 0.02$), with very high heterogeneity: $\text{Tau}^2 = 0.17$; $\chi^2 = 10.32$, df = 1 ($P = 0.001$); $I^2 = 90\%$. Both studies included in this subgroup were of lower methodological quality, and both were not age and BMI-matched in selecting cases and controls.

POF versus fertile controls versus postmenopausal controls (DHEA-S)

The POI/POF versus fertile controls subgroup showed a SMD (IV, random, 95% CI) −0.65 [−0.92, −0.37], with P value <0.05. POI/POF versus postmenopausal controls (n = 465) showed a SMD 1.12 [0.10, 2.14], $Z = 2.15$ ($P = 0.03$), heterogeneity: $\text{Tau}^2 = 0.77$; $\chi^2 = 35.54$, df = 2 ($P < 0.00001$); $I^2 = 94\%$, indicating that the DHEA-S levels are much reduced in POI/POF cases, but still show a higher concentration than DHEA-S in postmenopausal women. Test for subgroup differences showed $\chi^2 = 10.46$, df = 1 ($P = 0.001$), $I^2 = 90.4\%$, indicating substantial heterogeneity between these subgroups (Fig. 5).

Sensitivity analysis for DHEA-S concentrations

Sensitivity analysis was first conducted using the leave-one-out method by systematically omitting one study at a time, and the results were robust. Three studies (n = 210) that scored 6 or greater on the Newcastle-Ottawa Scale, considered as the best-quality studies, were chosen out of the eight studies, namely the studies by Doldi et al, Hartmann et al, and van der Stege et al. The study by van der Stege et al scored 7*s on the Newcastle-Ottawa Scale and was of the highest methodological quality. The studies by Doldi et al and Hartmann et al scored 6*s on the Newcastle-Ottawa Scale, even though the selection methodology for cases and controls was not clearly mentioned. The study by Doldi et al, was included here as we wanted to use at least three studies to conduct the sensitivity analysis and these three were the highest-quality studies comparing DHEA-S. These were the studies with the best methodological quality and considered having the least amount of bias; in all these studies cases, controls were age, and BMI-matched. This showed a SMD of −0.46 [−0.74, −0.17], $Z = 3.16$ ($P = 0.002$), with a drop in heterogeneity to 0%: heterogeneity: $\text{Tau}^2 = 0.00$; $\chi^2 = 1.43$, df = 2 ($P = 0.49$); $I^2 = 0\%$ (Fig. 6).

Androstenedione concentrations

Women suffering from POF have lower androstenedione in comparison with the controls according to data retrieved from 10 studies (n = 1,226) (Table 3), SMD (IV, random, 95% CI) −1.09 [−1.71, −0.48], with P value <0.05. There was significant heterogeneity observed in the meta-analysis ($I^2 = 95\%$). With heterogeneity: $\text{Tau}^2 = 0.92$; $\chi^2 = 180.64$, df = 9 ($P < 0.00001$); so, a random-effects model was used. The funnel plot was evaluated to check for publication bias. A subgroup analysis was done to verify the source of heterogeneity. The sensitively analysis showed a consistency among the studies (Fig. 7).

Smubgroup analysis for androstenedione concentrations

The subgroups were divided according to the various assays used for measurement. The first subgroup consisted of six studies that directly applied RIA for the measurement of DHEA-S concentrations. The second subgroup was the

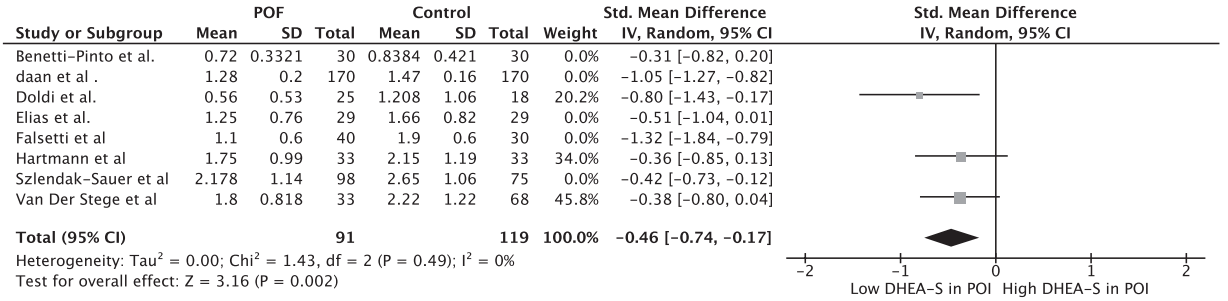

FIG. 6. Sensitivity analysis for DHEA-S concentrations in women with POI/POF compared with fertile controls. Three studies (n = 210) that scored >6 on the Newcastle-Ottawa Scale, considered as the best quality studies, were chosen out of the eight studies. DHEA-S, dehydroepiandrosterone sulfate; POF, primary ovarian failure; POI, premature ovarian insufficiency.

studies in which the assays used were either not mentioned or needed more clarification. This consisted of four studies (Fig., Supplemental Digital Content5, http://links.lww.com/MENO/A337).

Direct RIA subgroup

The subgroup consisted of six studies (n = 398). This subgroup showed a SMD of −1.10 [−1.61, −0.58], $Z = 4.17$ ($P < 0.0001$) with heterogeneity: $\text{Tau}^2 = 0.34$; $\chi^2 = 27.45$, df = 5 ($P < 0.0001$); $I^2 = 82\%$.The source of this heterogeneity could be explained by the fact that three of the six studies included in this particular subgroup happened to be studies with lower methodological quality. The study by Benetti-Pinto et al had no description of how cases and controls were selected; the study by Elias et al did not have age or BMI-matched cases and controls, nor did it have a description of how cases and controls were selected; and the study by Falsetti et al had no description of how cases and controls were selected, nor did it have a description whether both cases and controls used similar tests to check for exposure. We tried removing these studies from this subgroup and saw that the heterogeneity had substantially reduced to about 2%.

Other/nonspecified assay subgroup

The subgroup consisted of only four studies (n = 828). This subgroup showed a SMD of −1.08 [−2.30, 0.15], $Z = 1.73$ ($P = 0.08$) with very high heterogeneity: $\text{Tau}^2 = 1.52$; $\chi^2 = 150.25$, df = 3 ($P < 0.00001$); $I^2 = 98\%$. Two studies included in this subgroup were of lower methodological quality, and both were not age and BMI-matched in selecting cases and controls, namely the studies by Szlendak-Sauer et al and Daan et al; we tried removing these studies from this subgroup and saw that the heterogeneity had substantially reduced to about 5%.

POF versus fertile controls versus postmenopausal controls (androstenedione)

The POI/POF versus fertile controls subgroup showed a SMD (IV, random, 95% CI) −1.09 [−1.71, −0.48] with P value <0.05. POI/POF versus postmenopausal controls (n = 466) showed a SMD 0.27 [−0.13, 0.67], $Z = 1.34$ ($P = 0.18$), heterogeneity: $\text{Tau}^2 = 0.08$; $\chi^2 = 5.97$, df = 2 ($P = 0.05$); $I^2 = 66\%$. Test for subgroup differences showed $\chi^2 = 13.41$, df = 1 ($P = 0.0003$), $I^2 = 92.5\%$ (Fig., Supplemental Digital Content 6, http://links.lww.com/MENO/A338).

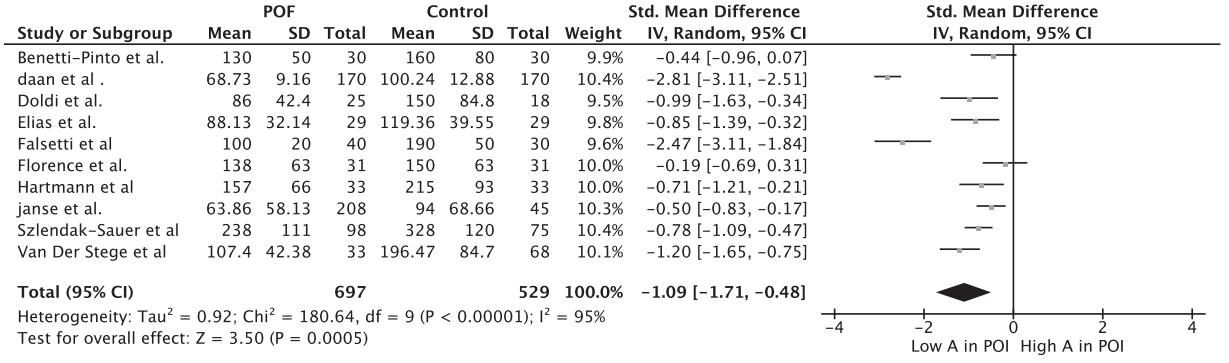

FIG. 7. Forest plot for meta-analysis of 10 comparative studies androstenedione concentrations in women with POI/POF compared with fertile controls. The values to the left of the solid line indicate SMD less than 1, decreased concentrations in POI/POF group. The values to the right of the solid line indicate SMD greater than 1, increased concentrations in POF/POI when compared to the control group. POF, primary ovarian failure; POI, premature ovarian insufficiency; SMD, standardized mean difference.

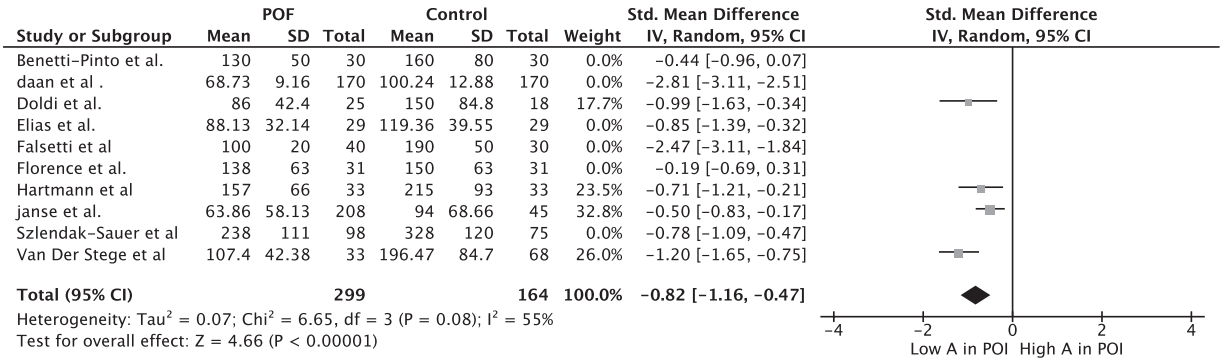

FIG. 8. Sensitivity analysis for androstenedione concentrations in women with POI/POF compared with fertile controls. Three studies (n = 463) that scored >6 on the Newcastle-Ottawa Scale, considered as the best quality studies, were chosen out of the 10 studies. POF, primary ovarian failure; POI, premature ovarian insufficiency.

Sensitivity analysis for androstenedione concentrations

Sensitivity analysis was first conducted using the leave-one-out method by systematically omitting one study at a time, and the results were robust. Then four studies (n = 463) that scored 6 or greater on the Newcastle-Ottawa Scale, considered as the best-quality studies, were chosen out of the 10 studies, namely the studies by Doldi et al, Hartmann et al, van der Stege et al, and Janse et al. The studies by van der Stege et al and Janse et al scored 7*s on the Newcastle-Ottawa Scale and were of the highest methodological quality. The studies by Doldi et al and Hartmann et al scored 6*s on the Newcastle-Ottawa Scale. This showed a SMD of −0.82 [−1.16, −0.47], $Z = 4.66$ ($P < 0.00001$), with a drop in heterogeneity to 55%: heterogeneity: $\text{Tau}^2 = 0.07$; $\chi^2 = 6.65$, df = 3 ($P = 0.08$); $I^2 = 55\%$ (Fig. 8).

Sex hormone-binding globulin concentrations

Women suffering from POI/POF did not seem to have a statistically significant difference to fertile controls with regards to SHBG levels according to data retrieved from nine studies (n = 1,613) (Table 4); SMD (IV, random, 95% CI) −0.13 [−0.35, 0.09] with *P* value >0.05 (not significant) (see Fig., Supplemental Digital Content 7, http://links.lww.com/MENO/A339).

DISCUSSION

In this systematic review and meta-analysis, we analyzed the various androgen concentrations including both the principal ovarian and adrenal androgens such as serum concentrations of total T, A, DHEA-S, and SHBG, and compared them with fertile controls and postmenopausal controls.

The pooled testosterone concentrations were found to be lower in women with POI/POF compared with fertile controls; SMD (IV, random, 95% CI) −0.73 [−0.99, −0.46] with *P* value <0.05. Testosterone concentrations in POI/POF versus postmenopausal controls were not statistically significant, with a SMD −0.05 [−0.68, 0.58], *P* value >0.05.

The pooled DHEA-S concentrations were found to be lower in women with POI/POF compared with fertile controls; SMD (IV, random, 95% CI) −0.65 [−0.92, −0.37] with *P* value <0.05. DHEA-S concentrations in POI/POF versus postmenopausal controls were statistically significant with a SMD 1.12 [0.10, 2.14], *P* value < 0.05, indicating that the DHEA-S levels are reduced in POI/POF, but still show a higher concentration than DHEA-S in postmenopausal controls. Women in the postmenopausal group are older than POI/POF women. It was only for DHEA-S that a significant decrease was noted in the postmenopausal group when comparing testosterone, DHEA-S, and androstenedione between POI/POF women and the postmenopausal group. The reasons for a significant decrease of DHEA-S seen in the postmenopausal group could be due to aging. The study by Labrie et al[44] has summarized that DHEA secretion will have already decreased by an average of 60% at time of menopause and will continue to decrease thereafter. We propose that this difference in decreasing pattern in serum DHEA-S with increasing age between POI/POF women and the postmenopausal group is mainly due to a higher compensatory mechanism in the relatively younger POI/POF women. In women of all ages, the inactive sex steroid precursor DHEA is mainly of adrenal origin. The study by Labrie et al[45,46] reported that the ovary is responsible for production of only 20% of circulating DHEA. The inactive DHEA will be transformed into the appropriate minute intracellular amounts of androgens to exert their physiological function. There is no biologically significant release into the serum for these intracellular androgens. This intracrine mechanism of androgen formation from DHEA could be responsible for why only for DHEA-S that a significant decrease was noted in the postmenopausal group, and why no significant difference was seen when comparing testosterone and androstenedione between POI/POF women and the postmenopausal group.

The pooled A concentrations were found to be lower in women with POI/POF compared with fertile controls; SMD (IV, random, 95% CI) −1.09 [−1.71, −0.48] with *P* value <0.05. Androstenedione (A) concentrations in POI/POF versus postmenopausal controls were again not statistically significant, with a SMD 0.27 [−0.13, 0.67], *P* value >0.05.

Women suffering from POI/POF did not seem to have a statistically significant difference compared to fertile controls with regards to SHBG levels according to data retrieved; SMD (IV, random, 95% CI) −0.13 [−0.35, 0.09], with P value >0.05.

There was significant heterogeneity between studies. Each of these androgens decreases with age. Most of the studies chosen had age-matched cases and controls, and only three studies did not match for age or BMI. So, we conducted a sensitivity analysis using only the best quality studies for each androgen assessed. In these studies, cases and controls were age and BMI-matched, and results showed substantial decrease in heterogeneity. Age and BMI were important factors in quality assessment for these studies involved. Some studies did not adjust for these factors. This could be one of the reasons for the heterogeneity between these studies. Another reason for heterogeneity between studies could be due to the differences in selecting controls and methodologies used for recruitment. Finally, the most important factor contributing towards the heterogeneity between studies could be explained by the fact that the studies used different assays for the measurement of androgen concentrations. This was apparently evident in the subgroup analysis using different assays (see Fig., Supplemental Digital Content3, 4, and 5, http://links.lww.com/MENO/A335, http://links.lww.com/MENO/A336, http://links.lww.com/MENO/A337). There was a substantial decrease in heterogeneity within most subgroups, which conducted analysis using the same assay. Few of the subgroups undertaken using the same assay showed reduced but still a significant level of heterogeneity. This could be explained by the fact that even though they were using the same kind of assay for measuring the concentrations, the intra and interassay coefficients used were widely variable. And, also, the source of this heterogeneity could be explained by the fact that few of these studies included happened to be lower-quality studies and were not age or BMI-matched or had no description of how cases and controls were chosen or if the same tests were used to ascertain exposure in cases and controls. The study by Benetti-Pinto et al[21] had no description of how cases and controls were selected. The study by Elias et al[7] did not have information on age or BMI for cases and controls, nor did it have a description of how cases and controls were selected. The study by Falsetti et al[38] had no description of how cases and controls were selected, nor did it have description whether both cases and controls used similar tests to check for exposure. The studies by Szlendak-Sauer et al[39] and Daan et al[43] were not age and BMI-matched in selecting cases and controls. When these studies with lower methodological quality were removed from the respective subgroups, the heterogeneity substantially reduced. So we conducted a sensitivity analysis using only studies with the best methodological quality, in which the most important confounding factors were age and BMI, and we observed significant results and substantial reduction in heterogeneity in all of the androgens assessed.

Taking all these factors contributing to heterogeneity into account, SMD was used to estimate the pooled concentrations and a random-effects model was used. SMD was preferred over a mean difference, because the methodologies used for estimation of the concentrations widely varied and this variation is already acknowledged by using SMD for estimation. Also, the sensitivity analysis using the best quality studies proved that the results obtained were robust. Therefore, we may interpret the results of this meta-analysis as being statistically significant.

The major androgens found in women in descending order of their serum concentrations include the following: DHEA-S, A, and T.[47] Of these androgens, T is the most potent.[48] It is secreted by the adrenal zonafasciculata (25%) and the ovarian stroma (25%), with the remaining 50% being produced from circulating A.[49] DHEA is a secretory product of the adrenal zonareticularis (50%) and the ovarian theca (20%), and 30% is derived from circulating DHEA-S, catalyzed by steroid sulphatase.[49] Androstenedione is secreted by the adrenal zonafasciculata (50%) and the ovarian stroma (50%).[49] We have identified decreased concentrations of these androgens, namely DHEA, which is produced mostly by the adrenal glands; however, it is still not clear whether this reduction is solely due to the ovarian component or if the adrenal component is also affected in idiopathic POI/POF.

The consequences of androgen deficiency have been discussed in very few studies. Androgens are thought to be one of the basic prerequisites for a healthy woman.[7-14] A lack of androgens may lead to symptoms of sexual dysfunction, such as decreased libido, loss of sexual responsiveness, or decreased sexual arousal.[7-14] Other clinical symptoms are a diminished sense of well-being, dysphoric mood, cognitive dysfunction, and persistent, unexplained loss of energy.[7-14]

The role of androgens in maintaining a woman's health has been receiving increasing attention, but still, there is considerable controversy relating to the role of androgens in women, and its clinical significance.[4,13-21] Testosterone at supraphysiological, but not at physiological, levels enhance the effectiveness of low-dose estrogen therapies at increasing women's sexual desire; however, the mechanism by which supraphysiological testosterone increases women's sexual desire in combination with an estrogen remains unknown.[12] Clinical evidence suggests that testosterone has anxiolytic and antidepressant benefits, with the potential to promote improved mood and mental health in women.[50] However, the neurobiological mechanisms underlying the protective effects of testosterone in males and females remain poorly understood. Selective androgens appear capable of improving early stages of folliculogenesis. Androgens, like T, appear effective in improving functional ovarian reserve in women with diminished ovarian reserve.[51] A study found that a low-testosterone status is a potentially important step in the development of POI/POF in women with endometriosis.[52] In women with diminished ovarian reserve, the basal T level presented a positive association with pregnancy outcome in in vitro fertilization.[52]

This meta-analysis has a few limitations. Very few studies were available for the topic of interest that would meet the eligibility criteria. Inclusion criteria were somewhat limited in databases assessed, some studies with lower methodological quality were included; this lead to significant heterogeneity between the studies. We have rectified these issues of including lower methodological quality studies by conducting a sensitivity analysis with higher methodological quality studies only. Each of these androgens decrease with age, most of the studies chosen had age-matched cases and controls. Only three studies did not match for age or BMI. Again, we conducted a sensitivity analysis using the best quality studies for each androgen assessed, in these studies cases and controls were age and BMI-matched; and the results seemed to be consistent and robust. Further, it is very difficult to measure and compare these sex steroids at lower ranges with accuracy using various assays. This was another challenge we faced while conducting this meta-analysis. Based on this limitation in assessment, we had to standardize the units of measurement and use a SMD and random-effects model to account for these differences. Currently, liquid chromatography-tandem mass spectrometry (LC-MS/MS) is a widely accepted and accurate assay for measuring low levels of steroids, such as androgens. Further comparative studies which employ LC-MS /MS, or similar assays, are needed to make sure the results obtained using various other assays can be depended upon.

Potential clinical value

Our meta-analysis has found that there is significant androgen deficiency in women with POI/POF. A significant reduction in androgen levels could lead to various complaints. Using androgen replacement therapy as an adjunct to pre-existing estrogen-based replacement therapy could be advantageous in relieving health complaints and promoting a more qualitative life in these women. Also, we suggest that, because the assays used for measurement of various androgens are widely variable, a standardized method for measurement needs to be developed for accurate assessment of these indices.

CONCLUSIONS

In conclusion, this meta-analysis demonstrated that women with POI/POF are at risk for decreased concentrations of total testosterone, DHEA-S, and androstenedione. SHBG concentrations did not show any statistical significance between POI/POF and controls. DHEA-S levels were more reduced in postmenopausal controls when compared with POI/POF cases. Testosterone and androstenedione concentrations did not show any statistical significance when POI/POF cases were compared with postmenopausal controls.

REFERENCES

1. Nelson LM. Primary ovarian insufficiency. *N Engl J Med* 2009;360:606-614.
2. De Vos M, Devroey P, Fauser B. Primary ovarian insufficiency. *Lancet* 2010;376:911-921.
3. Laven JSE. Primary ovarian insufficiency. *Semin Reproduct Med* 2016;34:230-234.
4. Webber L, Davies M, Anderson R, et al. ESHRE guideline: management of women with premature ovarian insufficiency. *Hum Reprod* 2016;31:926-937.
5. Welt CK. Primary ovarian insufficiency: a more accurate term for premature ovarian failure. *Clin Endocrinol* 2008;68:499-509.
6. Gleicher N, Kim A, Weghofer A, et al. Hypoandrogenism in association with diminished functional ovarian reserve. *Hum Reprod* 2013;28:1084-1091.
7. Elias AN, Pandian MR, Rojas FJ. Serum levels of androstenedione, testosterone and dehydroepiandrosterone sulfate in patients with premature ovarian failure to age-matched menstruating controls. *Gynecol Obstet Invest* 1997;43:47-48.
8. Bachmann G, Bancroft J, Braunstein G, et al. Female androgen insufficiency: the Princeton consensus statement on definition, classification, and assessment. *Fertil Steril* 2002;77:660-665.
9. Bachmann G, Oza D. Female androgen insufficiency. *Obstet Gynecol Clin N Am* 2006;33:589.
10. Apperloo MJA, van der Stege JG, Hoek A, Schultz W. In the mood for sex: the value of androgens. *J Sex Marital Ther* 2003;29:87-102.
11. Ryan J, Scali J, Carriere I, et al. Impact of a premature menopause on cognitive function in later life. *BJOG* 2014;121:1729-1739.
12. McHenry J, Carrier N, Hull E, Kabbaj M. Sex differences in anxiety and depression: role of testosterone. *Front Neuroendocrinol* 2014;35:42-57.
13. Davis SR, Wahlin-Jacobsen S. Testosterone in women: the clinical significance. *Lancet Diabetes Endocrinol* 2015;3:980-992.
14. Tyagi V, Scordo M, Yoon RS, Liporace FA, Greene LW. Revisiting the role of testosterone: are we missing something? *Rev Urol* 2017;19:16-24.
15. Bhasin S. Editorial: female androgen deficiency syndrome: an unproven hypothesis. *J Clin Endocrinol Metab* 2005;90:4970-4972.
16. Wierman ME, Arlt W, Basson R, et al. Androgen therapy in women: a reappraisal: an Endocrine Society Clinical Practice guideline. *J Clin Endocrinol Metab* 2014;99:3489-3510.
17. Davis S. Androgen replacement in women: a commentary. *J Clin Endocrinol Metab* 1999;84:1886-1891.
18. Guay A, Traish A. Testosterone therapy in women with androgen deficiency: its time has come. *Curr Opin Investig Drugs* 2010;11:1116-1126.
19. Mazer NA. Testosterone deficiency in women: etiologies, diagnosis, and emerging treatments. *Int J Fertil Womens Med* 2002;47:77-86.
20. Shohat-Tal A, Sen A, Barad DH, Kushnir V, Gleicher N. Genetics of androgen metabolism in women with infertility and hypoandrogenism. *Nat Rev Endocrinol* 2015;11:429-441.
21. Benetti-Pinto CL, Bedone AJ, Magna LA. Evaluation of serum androgen levels in women with premature ovarian failure. *Fertil Steril* 2005;83:508-510.
22. Mason M, Fonseca E, Ruiz JE, Moran C, Zarate A. Distribution of follicle-stimulating-hormone and luteinizing-hormone isoforms in sera from women with primary ovarian failure compared with that of normal reproductive and postmenopausal women. *Fertil Steril* 1992;58:60-65.
23. Nelson LM, Covington SN, Rebar RW. An update: spontaneous premature ovarian failure is not an early menopause. *Fertil Steril* 2005;83:1327-1332.
24. Hartmann BW, Kirchengast S, Albrecht A, Laml T, Soregi G, Huber JC. Androgen serum levels in women with premature ovarian failure compared to fertile and menopausal controls. *Gynecol Obstet Invest* 1997;44:127-131.
25. Bermudez JA, Moran C, Herrera J, Barahona E, Perez MC, Zarate A. Determination of the steroidogenic capacity in premature ovarian failure. *Fertil Steril* 1993;60:668-671.
26. Burger HG, Dudley EC, Cui JS, Dennerstein L, Hopper JL. A prospective longitudinal study of serum testosterone, dehydroepiandrosterone sulfate, and sex hormone-binding globulin levels through the menopause transition. *J Clin Endocrinol Metab* 2000;85:2832-2838.
27. Burger HG, Dudley EC, Robertson DM, Dennerstein L. Hormonal changes in the menopause transition. *Recent Prog Horm Res* 2002;57:257-275.
28. Davison SL, Bell R, Donath S, Montalto JG, Davis SR. Androgen levels in adult females: changes with age, menopause, and oophorectomy. *J Clin Endocrinol Metab* 2005;90:3847-3853.
29. Wells G. The Newcastle Ottawa Scale (NOS) for assessing the quality of non-randomised studies in meta-analysis. Proceedings of the Third Symposium on Systematic Reviews. Beyond the basics: improving quality and impact. Oxford, England; 2000.

30. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song, F. Methods for Meta-analysis in Medical Research. Chichester, United Kingdom: John Wiley & Sons; 2000.
31. Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration; 2014.
32. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-560.
33. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-634.
34. Kalantaridou SN, Calis KA, Vanderhoof VH, et al. Testosterone deficiency in young women with 46,XX spontaneous premature ovarian failure. *Fertil Steril* 2006;86:1475-1482.
35. van der Stege JG, Groen H, van Zadelhoff SJ, et al. Decreased androgen concentrations and diminished general and sexual well-being in women with premature ovarian failure. *Menopause* 2008;15:23-31.
36. Gulhan I, Bozkaya G, Uyar I, Oztekin D, Pamuk BO, Dogan E. Serum lipid levels in women with premature ovarian failure. *Menopause* 2012;19:1231-1234.
37. Doldi N, Belvisi L, Bassan M, Fusi FM, Ferrari A. Premature ovarian failure: steroid synthesis and autoimmunity. *Gynecol Endocrinol* 1998;12:23-28.
38. Falsetti L, Scalchi S, Villani MT, Bugari G. Premature ovarian failure. *Gynecol Endocrinol* 1999;13:189-195.
39. Szlendak-Sauer K, Jakubik D, Kunicki M, Skórska J, Smolarczyk R. Autoimmune polyglandular syndrome type 3 (APS-3) among patients with premature ovarian insufficiency (POI). *Eur J Obstet Gynecol Reprod Biol* 2016;203:61-65.
40. Ates S, Yesil G, Sevket O, Molla T, Yildiz S. Comparison of metabolic profile and abdominal fat distribution between karyotypically normal women with premature ovarian insufficiency and age matched controls. *Maturitas* 2014;79:306-310.
41. Janse F, Eijkemans MJ, Goverde AJ, et al. Assessment of androgen concentration in women: liquid chromatography-tandem mass spectrometry and extraction RIA show comparable results. *Eur J Endocrinol* 2011;165:925-933.
42. Omu FE, Elbiaa A, Ghafour A, Gadalla I, Omu AE. Beneficial effects of tibolone on sexual dysfunction in women with premature ovarian failure (POF). *Health* 2016;8:857-867.
43. Daan NMP, Jaspers L, Koster MPH, et al. Androgen levels in women with various forms of ovarian dysfunction: associations with cardiometabolic features. *Hum Reprod* 2015;20:2376-2386.
44. Labrie F, Martel C, Balser J. Intracrine formation from dehydroepiandrosterone versus ovarian secretion of testosterone response. *Menopause* 2011;18:587-590.
45. Labrie F, Martel C, Balser J. Wide distribution of the serum dehydroepiandrosterone and sex steroid levels in postmenopausal women: role of the ovary? *Menopause* 2011;18:30-43.
46. Labrie F, Martel C, Belanger A, Pelletier G. Androgens in women are essentially made from DHEA in each peripheral tissue according to intracrinology. *J Steroid Biochem Mol Biol* 2017;168:9-18.
47. Rivera-Woll LM, Papalia M, Davis SR, Burger HG. Androgen insufficiency in women: diagnostic and therapeutic implications. *Hum Reprod Update* 2004;10:421-432.
48. Burger HG. Androgen production in women. *Fertil Steril* 2002;77 (Suppl 4): S3-S5.
49. Cappelletti M, Wallen K. Increasing women's sexual desire: the comparative effectiveness of estrogens and androgens. *Horm Behav* 2016;78:178-193.
50. Gleicher N, Weghofer A, Barad DH. The role of androgens in follicle maturation and ovulation induction: friend or foe of infertility treatment? *Reprod Biol Endocrinol* 2011;9:116.
51. Ono YJ, Tanabe A, Nakamura Y, et al. A low-testosterone state associated with endometrioma leads to the apoptosis of granulosa cells. *PLoS One* 2014;9:e11568.
52. Lu Q, Shen H, Li Y, et al. Low testosterone levels in women with diminished ovarian reserve impair embryo implantation rate: a retrospective case-control study. *J Assist Reprod Genet* 2014;31:485-491.