PEER REVIEW HISTORY

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

(This paper received three reviews from its previous journal but only two reviewers agreed to published their review.)

ARTICLE DETAILS

TITLE (PROVISIONAL)	Acupuncture or metformin to improve insulin resistance in women with polycystic ovary syndrome: Study protocol of a combined multinational cross sectional case-control study and a randomized controlled trial
AUTHORS	Stener-Victorin, Elisabet; Zhang, Haolin; Rong, LI; Friden, Cecilia; Li, Dong; Wang, Wei; Wang, Haining; Chang, Cuiqing; Li, Shi; Huo, ZeJun; Zhang, Hua; Ji, Xiaolan; Linden-Hirschberg, Angelica; Jie, Qiao

VERSION 1 – REVIEW

REVIEWER	Renato Pasquali University Alma Mater Studiorum, Bologna, Italy
REVIEW RETURNED	04-Jul-2018

GENERAL COMMENTS	This paper describes a prospective study (RCT) aimed at assessing the impact of acupuncture associated with lifestyle intervention vs. metformin + lifestyle intervention on insulin resistance (defined by HOMA-IR) and HbA1c in women with PCOS with overweight or obesity (recruited from Sweden and China). Undoubtedly, the project is of interest, considering the data published by ES-V on the effects of acupuncture. Furthermore, the possibility of collecting tissues and whole blood specimen is a quality factor added to the project Specific comments: • Please explain why 114 women were enrolled in relation to the out as HbA1c and 303 in relation to the HOMA-IR outcome • It is not clear to me whether the assessment of ovarian morphology is standardized and the technology is the same • Since a slow-release formulation of metformin is also available, why not use this that has, however, much less collateral effects (gastrointestinal, etc) ?? • I would like to know why an endometrial biopsy is proposed • What methodologies will be used for the measurement of NE-A-DA? • On what basis was the treatment planned in 4 months?

REVIEWER	Florent Besnier Institute of Cardiovascular and Metabolic Diseases, Toulouse, France.
REVIEW RETURNED	12-Jul-2018

GENERAL COMMENTS	This bicentric clinical trial proposed by the Swedish and Chinese teams of Elisabet Stener-Victorin and Jie Qiao aims to compare in a Randomized Control Trial the effectiveness of 3 modalities of interventions to improve insulin sensitivity in women with polycystic ovary Syndrome (POCS). Groups are: 1/electro-acupuncture with lifestyle management 2/metformin with lifestyle management 3/lifestyle management alone The study also aim to compare in a case-control cross selectional study (healthy women vs POCS women) glycaemic profile. Nowadays, and for most chronic diseases such as diabetes, obesity ... lifestyle modifications must be encouraged and must be the norm even in clinical trials (understand: the reference group) when firsts outcomes are in line with glycaemic control. One of the strengths of this protocol study is that the control group is not a sedentary control group. It will benefit from dietary and physical activity advices. The other two groups are comparative groups. Of course, this influences the sample size of the study, which must be larger. However, the impact is beneficial mainly in terms of public health and for all participants (including for non-responders to acupuncture). Furthermore, fields of application are multiple and extrapolated in various pathologies of glucose metabolism. It should also be noted that the team has already experimented the acupuncture technique, which offers both acute and chronic effects possibly mediated by the autonomic nervous system (Kokosar M et al. Sci Rep 2018; Stener-Victorin E et al. Obes Sci Pract. 2016). To facilitate the reviewing process and the lecturer of the document, some of major precision in the design (main hypothesis & objective are confused) and a clearly organization of the paper are needed because some of informations are detailed in different parts of the manuscript. First, numbered each line continues not for each page. Title: The meaning of the title is fully understandable but maybe it misses a word in the first part. I propose: Acupuncture or metformin to improve insulin resistance… or to treat IR …. or to improve glycaemic control … I have a doubt with this formulation : “Acupuncture or metformin for insulin resistance”. Abstract line 13: you not designed your study with metformin and electroA alone. Abstract line 22 to 25: please numbered your groups in line with the main text. Abstract, Methods section: you should precise in the study design if this is a non-inferiority / superiority / or equivalence design. That is also not clear in the main text. Abstract: Ethics and dissemination: “The study is performed according to good clinical practice and conducted in accordance with the Declaration of Helsinki”. Remove this sentence, your study is already register in ClinicalTrial.gov. in accordance with ICJME recommendations. Add the expected results. And may the

	expected benefits for a larger population that could benefit from acupuncture? Key Words: Metformin is cited two time. Please remove one of them. Background line 46: before reference number 15 . You should add some of the limits of this pilot study. For example, you could explain that your study group was not controlled. That would clarified for the lecturer how this new RCT study is original and methodologically relevant. Line 3 page 4: you stated that "We hypothesize that acupuncture is equally effective as metformin (both treatments combined with lifestyle management) in improving whole body glucose homeostasis in insulin resistant women with PCOS". This is your main hypothesis? I understand that your study is an equivalence study design. Which implies giving the upper and lower clinical limits of HOMA and HbA1C to validate the results. Please clarified on the methodology part. Sample size is slight different if you are testing equivalence or non-inferiority or superiority design. Secondly: "….and that both are superior to lifestyle management alone" This is one of your secondary hypothesis? Please clarified Background page 6 line 52 to page 8 line 10. This part is confused and mix hypothesis and objectives and main outcomes. Please clarified your main and second hypothesis; then your main and second objectives (in line with hypothesis) and your main outcome. Line 17 page 8: the first part of your study (Case-Control) is not prospective. Cross-selectional study? Design, page 8: you should add the randomization and treatment allocation in this part. Please, add the block size. (For example and if applicable: randomization by block with equal block size = 3, and balanced allocation ratio 1:1:1) Participants: In Inclusion criteria: "Age 18 to 40 years". You should precise >18 and ≤40 or <41. In Exclusion criteria: "age > 40" should be remove. You precise it in your inclusion criteria. Exclusion criteria 9: the average age for participant is between 18 to 40. Do you measures the level of hCG hormone, pre and post intervention? It is not indicated in the main text and in the table 2. Inclusion/Exclusion criteria for Control and PCOS women: your first outcomes is HbA1c and or HOMA-IR. Physical activity plays a major role in the evolution of these variables and it would be desirable to add in your criteria whether or not you include physically active people. And what is your cut-off for both "inactivity" and for "sedentary" because of different lifestyle (hour/week?). (For example see González K, and al. Physical Inactivity, Sedentary Behaviour and Chronic Diseases. 2017 doi:10.4082/kjfm.2017.38.3.111). The length of sedentary is also important. The topic of Physical activity may be the main confounding factor of your project according to your first outcomes and need to be documented. This could be an important bias at the end of the study. Life style management part: To my point of view, this part do not give enough information. Physical activity is one of the most factor that would influence your main outcome (HbA1c/HOMA-IR). You inform the lecturer that participant will count the number of steps. How are they recorded? Is there a goal to reach each day? (WHO recommends 10.000 steps/day). You are not mentioned the Physical activity guidelines for weight loss in overweight/obese

	people (your population). How do you take into account women who do cycling or gymnastic or dance (others activities than walking)? How do you take into account daily physical activity? The intensity and duration of each session of physical activity and the overall volume of exercise have a major impact on glycaemic control and have to be describe. Can you use accelerometers or smartphone app? It may be necessary to convert the physical activity done each day and each week into "calories". So at the end of the study, you can compare the number of calories spent in each group, regardless of the nature of the activity (walking, or cycling, or gym ...). This part is crucial. At the end of the study each group have to spend the same total energy expenditure and you have to describe this topic. Furthermore, it is well known than one session of exercise could improve muscle glucose metabolism during more than 24h. This acute effect (mostly mediated by GLUT4 protein) is not take into account for baseline and post intervention blood test. A resting washout period is needed between the last session of exercise and the blood test to guaranty glycaemic index in standard condition. Electro acupuncture: this part is very well detailed. Are there non-responders to this technique? Primary outcome page 14 line 24: you are mixing 2 outcomes with two groups. PCOS vs Control at baseline and PCOS women randomized in the 3 groups with post vs pre evaluations after 4 months of the intervention. Please clarified your design. In this part, you could remove line 33 to 35. Secondary outcome part: list of the outcome only with units. Remove all sentences as “all women with be examined by DXA to measure lean and fat mass and bone mineral density using a Lunar Prodigy Advance whole body scanner (GE Medical Systems)”. This sentence have to be be add to “Study procedure” part. Sample size: based on HbA1c, my sample size calculation is different than 23. With a mean difference of 0.86 for HbA1c and a SD of 1.4 = 42 for each group. Line 32 page 17: number of subject=23 subjects/groups as mentioned in “sample size” is not equal to 116. Statistical analyses: “when the intended number of participants for Hba1c have been reached, an interim analyses will be performed”. Repeated Significance Tests with interim analysis according to Haybittle–Peto boundary is a rule for deciding when to stop a clinical trial prematurely. I appreciate this transparency statement by the authors. Interim analyses are statistically complex to implement but methodologically necessary. Who is performing this interim analyses? Classical interim analyses are performed by an independent statistician who should be a person other than the regular study statistician. Line 38 page 17: “The stop criterion are meant for both co-primary and covers two group comparisons”. Witch two groups are you talking about? Acupuncture and Metformin group? Are you stopping the research if these two groups (Acup. vs Metf.) are equivalent in the HbA1c evolution according to one of your hypothesis? Or if one of these two groups (or both) will show a superiority vs Lifestyle group? This two hypothesis are different and not clear. Line 43 page 17: May I have missed this information but “posthoc corrections as given below.” are not mentioned below. Is it Bonferroni correction with 0.05/(number of groups) = 0.0167?

	Trial Status part: the dates mentioned here are 2015, 2016 and 2019, with a gap between them. It could be informative to add the number of women already recruited in the middle of 2018 in each country. How about women assessed for eligibility but not randomized because of declined to participate/pregnant … etc? A CONSORT flow diagram will be necessary for the main publication and should mentioned here.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Renato Pasquali
Institution and Country: University Alma Mater Studiorum, Bologna, Italy Please state any competing interests or state 'None declared': No competing interest

This paper describes a prospective study (RCT) aimed at assessing the impact of acupuncture associated with lifestyle intervention vs. metformin + lifestyle intervention on insulin resistance (defined by HOMA-IR) and HbA1c in women with PCOS with overweight or obesity (recruited from Sweden and China).
Undoubtedly, the project is of interest, considering the data published by ES-V on the effects of acupuncture. Furthermore, the possibility of collecting tissues and whole blood specimen is a quality factor added to the project

Specific comments:
• Please explain why 114 women were enrolled in relation to the out as HbA1c and 303 in relation to the HOMA-IR outcome
Response: After consulting a new statistical expert we realize that our sample size calculation is not entirely correct. We have therefore re-calculated and decided to have only one primary outcome variable = Hba1c, see page 14, Line 427-443. The reason for this is that we were advised not to do an interim analyses as that is primarily to be used if you need to terminate the experiment due to detrimental effects. Below is a detailed description including assumptions and calculations.

<u>Power calculation - HbA1c change between 0 to 4 months</u>
Assumptions

- Change between 4 and 6 months is negligible and we can use results/effect difference from 6 months in the calculations. Six months of acupuncture was given in an uncontrolled trial (unpublished, see study protocol: Zheng Y et al. BMJ Open 2015; 5:e007757).
- The control group (lifestyle management only) has on average no change in HbA1c.
- All three groups have the same variation/standard deviation.

Sample size calculation

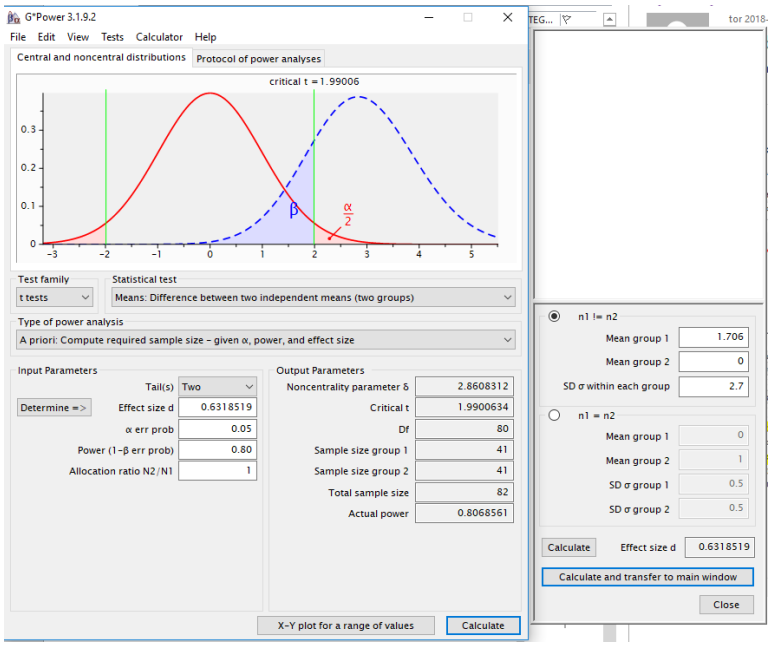
Calculations are based on t-test between two groups. This is due to the fact that it is the pairwise comparisons that are of main interest (not overall F-test/ANOVA).

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Hba1c 6 mån	5.2760	75	.29585	.03416
	Hba1c 0 mån	5.4413	75	.38171	.04408
Pair 2	Hba1c 6 mån mmol	44.6667	75	3.04619	.35174
	Hba1c 0 mån mmol	46.3733	75	3.83361	.44267

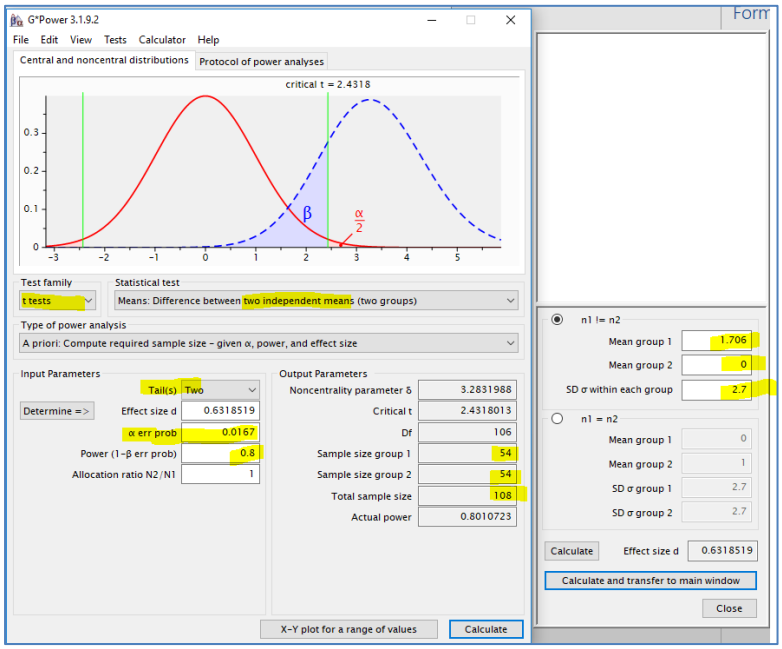
Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Hba1c 6 mån - Hba1c 0 mån	-.16533	.26432	.03052	-.22615	-.10452	-5.417	74	.000
Pair 2	Hba1c 6 mån mmol - Hba1c 0 mån mmol	-1.70667	2.73008	.31524	-2.33480	-1.07853	-5.414	74	.000



The result show that 41 * 3 = 123 is enough to prove a difference between active and control group (repeated pairwise t-test) on -1.7 unites (effect size 1.7/2.7=0.63) with 80 % power (significance, p = 0.05, unadjusted pairwise comparisons).

Since the pairs of tests will be repeated between all three groups, the significance level can be adjusted to 0.05/3 = 0.0167. Below we give samples size calculation with this approach.



The result show that 54 * 3 = 162 is enough to prove a difference between active and control group (repeated pairwise t-test) on -1.7 unites (effect size 1.7/2.7=0.63) with 80 % power (significance, p = 0.0167). However, this can be considered extremely conservative and we therefore decided to ignore the Bonferroni correction approach.

• It is not clear to me whether the assessment of ovarian morphology is standardized and the technology is the same

Response: The ovarian morphology is standardized and follow the Rotterdam criteria.
In addition to measure of ovarian volume and number of antral follicle <9 mm, we also measure size of the biggest follicle, peripheral localisation (yes/no) and estimate of increased

GYNECOLOGICAL EXAMINATION

	Y	N		Y	N
VVP normal	☐	☐			
Uterus palpation normal	☐	☐			
Vaginal ulj			**Endometrial thickness:** ☐ mm		
Right ovary			Left ovary		
Length:	☐			☐	
AP	☐			☐	
Transverse:	☐			☐	
Nr of antral follicles <9 mm	☐			☐	
Size of biggest follicle, cm^3	☐			☐	
Peripheral localisation	☐	☐		☐	☐
Increased stroma	☐	☐		☐	☐

stroma (yes/no) as well as endometrial thickness. Please see protocol below.

• Since a slow-release formulation of metformin is also available, why not use this that has, however, much less collateral effects (gastrointestinal, etc)??
Response: The reason for not using the slow-release formulation of metformin is because it is not available in Sweden and we thought it was important to use the standard formulation of metformin as one of the treatment groups.

• I would like to know why an endometrial biopsy is proposed
Response: Endometrial biopsies are proposed to further investigate molecular function and how it is affected by the interventions. Hirschberg et al has previous investigated the effect of lifestyle intervention on sex steroid receptor expression as well as insulin signalling molecules (see references below). Here we will deepen these analyses as well as we will have the opportunity to investigate cross-talk between skeletal muscle and adipose tissue.

Progesterone Receptors and Proliferation of the **Endometrium** in Obese Women With Polycystic Ovary Syndrome-A Lifestyle Intervention Study. Paulson M, Sahlin L, **Hirschberg** AL. J Clin Endocrinol Metab. 2017 Apr 1;102(4):1244-1253. doi: 10.1210/jc.2016-3155. PMID: 28388727

Endometrial Expression of Estrogen Receptors and the Androgen Receptor in Women With Polycystic Ovary Syndrome: A Lifestyle Intervention Study. Hulchiy M, Nybacka Å, Sahlin L, **Hirschberg** AL. J Clin Endocrinol Metab. 2016 Feb;101(2):561-71. doi: 10.1210/jc.2015-3803. Epub 2015 Dec 9. PMID: 26649621

Lifestyle intervention up-regulates gene and protein levels of molecules involved in insulin signaling in the **endometrium** of overweight/obese women with polycystic ovary syndrome. Ujvari D, Hulchiy M, Calaby A, Nybacka Å, Byström B, **Hirschberg** AL. Hum Reprod. 2014 Jul;29(7):1526-35. Epub 2014 May 19. PMID: 24842895

• What methodologies will be used for the measurement of NE-A-DA?
Response: Catecholamine's and metabolites (noradrenaline, dopamine, DOPAC, HVA, serotonin, and 5HIAA) will be analysed on a split-fraction HPLC-ED system (Prieto-Garcia L, et al. Psychoneuroendocrinology 2015; 62:392-402). This has been added to the material and methods, Page 7, Line 214. Of note, mentioned hormones, adipokines and lipids are examples of what to be analysed.

• On what basis was the treatment planned in 4 months?
Response: We have previously done a pilot study of 5 weeks treatment with acupuncture **(**Stener-Victorin E, et al Obes Sci Pract 2016; 2:426-435). In collaboration with colleagues in China we have done one pilot study with same treatment protocol for 6 months (unpublished, see study protocol: Zheng Y et al. BMJ Open 2015; 5:e007757). The experience is that the compliance for 6 months is not 100%. As we also investigate the effect of metformin which need at least 3 months to have an optimal effect we decided to go for 4 months (16 weeks) of treatment.

Reviewer: 2
Reviewer Name: Florent Besnier
Institution and Country: Institute of Cardiovascular and Metabolic Diseases, Toulouse, France.
Please state any competing interests or state 'None declared': None declared

This bicentric clinical trial proposed by the Swedish and Chinese teams of Elisabet Stener-Victorin and Jie Qiao aims to compare in a Randomized Control Trial the effectiveness of 3 modalities of interventions to improve insulin sensitivity in women with polycystic ovary Syndrome (POCS).
Groups are:
1/electro-acupuncture with lifestyle management 2/metformin with lifestyle management 3/lifestyle management alone
The study also aim to compare in a case-control cross selectional study (healthy women vs POCS women) glycaemic profile.

Nowadays, and for most chronic diseases such as diabetes, obesity ... lifestyle modifications must be encouraged and must be the norm even in clinical trials (understand: the reference group) when firsts outcomes are in line with glycaemic control. One of the strengths of this protocol study is that the control group is not a sedentary control group. It will benefit from dietary and physical activity advices. The other two groups are comparative groups. Of course, this influences the sample size of the study, which must be larger. However, the impact is beneficial mainly in terms of public health and for all participants (including for non-responders to acupuncture). Furthermore, fields of application are multiple and extrapolated in various pathologies of glucose metabolism. It should also be noted that the team has already experimented the acupuncture technique, which offers both acute and chronic effects possibly mediated by the autonomic nervous system (Kokosar M et al. Sci Rep 2018; Stener-Victorin E et al. Obes Sci Pract. 2016).
To facilitate the reviewing process and the lecturer of the document, some of major precision in the design (main hypothesis & objective are confused) and a clearly organization of the paper are needed because some of information are detailed in different parts of the manuscript.

First, numbered each line continues not for each page.
Response: This has been done.

Title: The meaning of the title is fully understandable but maybe it misses a word in the first part. I propose: Acupuncture or metformin to improve insulin resistance… or to treat IR …. or to improve glycaemic control … I have a doubt with this formulation : "Acupuncture or metformin for insulin resistance".

Response: We agree and changed to “Acupuncture or metformin to improve insulin resistance”.

Abstract line 13: you not designed your study with metformin and electroA alone.
Abstract line 22 to 25: please numbered your groups in line with the main text.
Abstract, Methods section: you should precise in the study design if this is a non-inferiority / superiority / or equivalence design. That is also not clear in the main text.
Abstract: Ethics and dissemination: “The study is performed according to good clinical practice and conducted in accordance with the Declaration of Helsinki”. Remove this sentence, your study is already register in ClinicalTrial.gov. in accordance with ICJME recommendations. Add the expected results. And may the expected benefits for a larger population that could benefit from acupuncture?
Response: All suggested changes has been done. Please see highlighted version of the manuscript, page 2, line 40-68.

Key Words: Metformin is cited two time. Please remove one of them.
Response: Done.

Background line 46: before reference number 15. You should add some of the limits of this pilot study. For example, you could explain that your study group was not controlled. That would clarified for the lecturer how this new RCT study is original and methodologically relevant.
Response: As stated, in a pilot study we have shown that 5 weeks of EA improve glycemic control in women with PCOS. Of note, this was an uncontrolled trial and it is therefore of importance to compare acupuncture with first line treatment, lifestyle management to investigate the effectiveness. See page 4, line 116-117.

Line 3 page 4: you stated that “We hypothesize that acupuncture is equally effective as metformin (both treatments combined with lifestyle management) in improving whole body glucose homeostasis in insulin resistant women with PCOS”. This is your main hypothesis? I understand that your study is an equivalence study design. Which implies giving the upper and lower clinical limits of HOMA and HbA1C to validate the results.
Please clarified on the methodology part. Sample size is slight different if you are testing equivalence or non-inferiority or superiority design.
Secondly: “….and that both are superior to lifestyle management alone” This is one of your secondary hypothesis?
Response: We agree that it was not clear and even incorrect as written. As given above in the response to Reviewer #1, we have performed new samples size calculation. Further, this is not an equivalence study, it is a superiority design. All statements that acupuncture and metformin are equally effective has been removed from the manuscript.
Throughout the ms this has been corrected. Please see a complete response to all questions below.
Sample size in the mansucript, see page 14, Line 427-443.

Power calculation - HbA1c change between 0 to 4 months

Assumptions

- Change between 4 and 6 months is negligible and we can use results/effect difference from 6 months in the calculations.
- The control group (lifestyle management only) has on average no change in HbA1c.
- All three groups have the same variation/standard deviation.

Sample size calculation

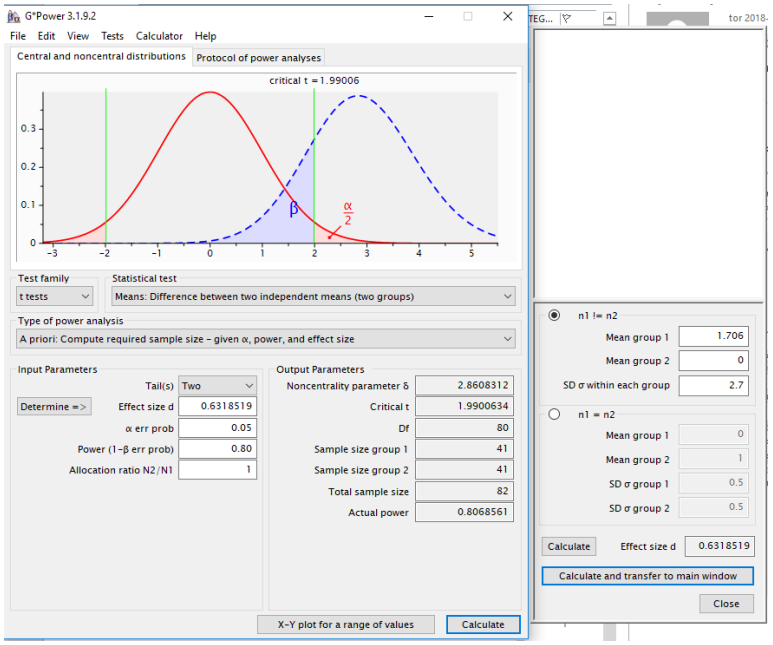
Calculations are based on t-test between two groups. This is due to the fact that it is the pairwise comparisons that are of main interest (not overall F-test/ANOVA).

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Hba1c 6 mån	5.2760	75	.29585	.03416
	Hba1c 0 mån	5.4413	75	.38171	.04408
Pair 2	Hba1c 6 mån mmol	44.6667	75	3.04619	.35174
	Hba1c 0 mån mmol	46.3733	75	3.83361	.44267

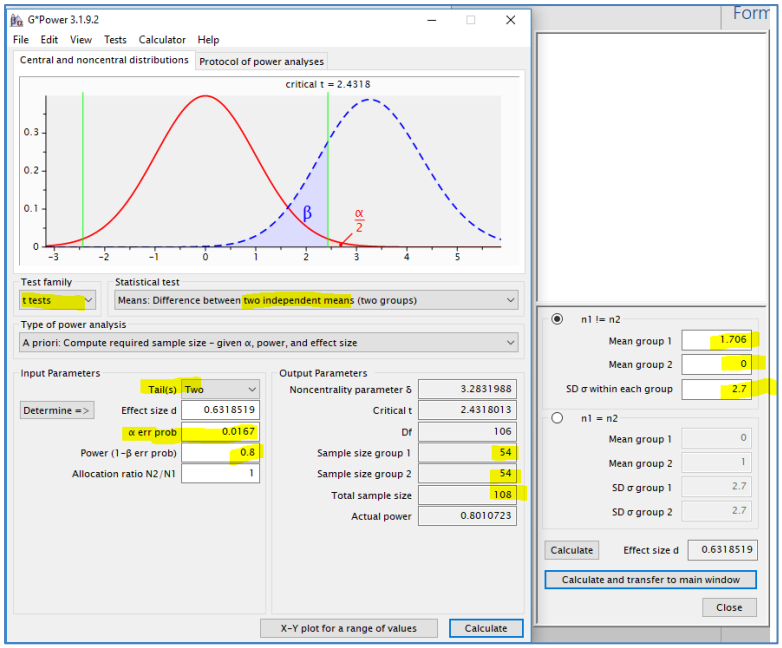
Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Hba1c 6 mån - Hba1c 0 mån	-.16533	.26432	.03052	-.22615	-.10452	-5.417	74	.000
Pair 2	Hba1c 6 mån mmol - Hba1c 0 mån mmol	-1.70667	2.73008	.31524	-2.33480	-1.07853	-5.414	74	.000



The result show that 41 * 3 = 123 is enough to prove a difference between active and control group (repeated pairwise t-test) on -1.7 unites (effect size 1.7/2.7=0.63) with 80 % power (significance, p = 0,05, unadjusted pairwise comparisons).

Since the pairs of tests will be repeated between all three groups, the significance level can be adjusted to 0.05/3 = 0.0167. However, this can be considered extremely conservative and we therefore decided not to do the Bonferroni correction.
Below we give samples size calculation with this approach.

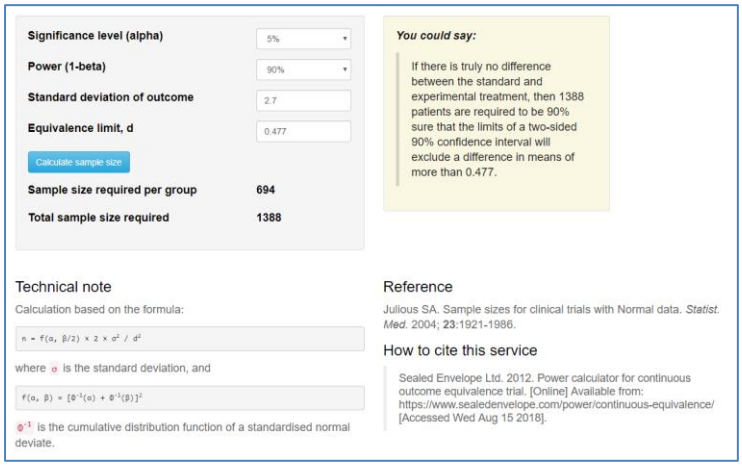


The result show that 54 * 3 = 162 is enough to prove a difference between active and control group (repeated pairwise t-test) on -1.7 unites (effect size 1.7/2.7=0.63) with 80 % power (significance, p = 0.0167).

Equivalence trial

In order to find that the two active treatments are equivalent, a so-called "equivalence test" is made. It is then important to classify at what limit / difference a treatment ceases to be clinically equivalent to the other. In Equivalence Test, it is the "type 2 error" (ie not rejecting an incorrect zero hypothesis) which is the serious error, therefore power is set to 90% and the significance level is allowed to be 10%.

In an example, given that SD = 2.7, the effect size for a four month change in HbA1c is then 1.71 / 2.73 = 0.626. A classic limit for a high effect size is 0.8, which would correspond to a difference of 0.8 * 2.73 = 2.18. If any of the two active treatments differ more than 2.18 - 1.71 = 0.477 from the other, then they may no longer be considered equivalent.



The above calculation shows that 694 individuals in the respective active treatment group (1388 in total) would be required to prove equivalence between treatments (when a maximum difference of 0.477 is maximum).

If the difference between the two active groups is allowed to be max half, that is, one group has a decrease of -1.7 / 2 = -0.85 units less over the period (which is a lot) then it would still be required at least 219 * 2 = 438 people to prove the equivalent of 90% power.

All statements that acupuncture and metformin are equally effective has been removed from the manuscript.

Please clarified Background page 6 line 52 to page 8 line 10. This part is confused and mix hypothesis and objectives and main outcomes. Please clarified your main and second hypothesis; then your main and second objectives (in line with hypothesis) and your main outcome.
Response: Apologize for the confusion. Hypotheses are now presented as main and secondary, followed by study design and randomization, objectives and outcome measures. See Page 5 to 9.

Line 17 page 8: the first part of your study (Case-Control) is not prospective. Cross-selectional study?
Response: Agree, this has been changed in abstract and methods.

Design, page 8: you should add the randomization and treatment allocation in this part. Please, add the block size. (For example and if applicable: randomization by block with equal block size = 3, and balanced allocation ratio 1:1:1)
Response: "Randomization and treatment allocation" has been moved to immediately after "Study Design".
We have also included the block size, see page 5, line 150-156: The randomization is stratified across the factors age and BMI and separated by study site with a balanced allocation ratios 1:1:1. Randomization is performed in blocks with a variable block size between 3 and 15; e.g. First there is a block of 12, when it is full it is followed by a block of 9 and thereafter a block of 3. The order of the block sizes are unknown to the participating study sites and also differs among the strata's.

Participants:
In Inclusion criteria: "Age 18 to 40 years". You should precise >18 and ≤40 or <41.
Response: Se changes Page 9, line 265: Age ≥18 to ≤40 years.

In Exclusion criteria: "age > 40" should be remove. You precise it in your inclusion criteria.
Response: Done.

Exclusion criteria 9: the average age for participant is between 18 to 40. Do you measures the level of hCG hormone, pre and post intervention? It is not indicated in the main text and in the table 2.
Response: Yes it is already given in the main text, page 14, line 421-422: "Women with PCOS who are randomized are informed that they should use contraception that are non-hormonal."

Inclusion/Exclusion criteria for Control and PCOS women: your first outcomes is HbA1c and or HOMA-IR. Physical activity plays a major role in the evolution of these variables and it would be desirable to add in your criteria whether or not you include physically active people. And what is your cut-off for both "inactivity" and for "sedentary" because of different lifestyle (hour/week?). (For example see González K, and al. Physical Inactivity, Sedentary Behaviour and Chronic Diseases. 2017 doi:10.4082/kjfm.2017.38.3.111). The length of sedentary is also important. The topic of Physical activity may be the main confounding factor of your project according to your first outcomes and need to be documented. This could be an important bias at the end of the study.
Response: We do agree that the amount of physical activity may influence most of our outcomes. Of note, we have two outcome measures evaluating the amount of physical activity. 1) IPAQ questionnaire and 2) number of steps per week is reported every week by sms and every 4 week by a phone call. A woman who do daily exercise and are competing are likely not interested to participate in this trial.

Life style management part: To my point of view, this part do not give enough information. Physical activity is one of the most factor that would influence your main outcome (HbA1c/HOMA-IR).
You inform the lecturer that participant will count the number of steps. How are they recorded? Is there a goal to reach each day? (WHO recommends 10.000 steps/day). You are not mentioned the Physical activity guidelines for weight loss in overweight/obese people (your population). How do you take into account women who do cycling or gymnastic or dance (others activities than walking)? How do you take into account daily physical activity? The intensity and duration of each session of physical activity and the overall volume of exercise have a major impact on glycaemic control and have to be describe. Can you use accelerometers or smartphone app? It may be necessary to convert the physical activity done each day and each week into "calories". So at the end of the study, you can

compare the number of calories spent in each group, regardless of the nature of the activity (walking, or cycling, or gym ...). This part is crucial. At the end of the study each group have to spend the same total energy expenditure and you have to describe this topic.
Furthermore, it is well known than one session of exercise could improve muscle glucose metabolism during more than 24h. This acute effect (mostly mediated by GLUT4 protein) is not take into account for baseline and post intervention blood test. A resting washout period is needed between the last session of exercise and the blood test to guaranty glycaemic index in standard condition.
Response: As given above, we agree that the amount of physical activity may influence the outcome. With IPAQ we will be able to calculate METs and with reporting of number of steps per week we will get information about degree of activity. Given that it's a RCT we assume that the distribution between sedentary participants and slightly more active participants are equally distributed. Participants are told not to do any exercise, smoke or drink alcohol 24 hour prior each testing (baseline, after 4 months of treatment and at follow up 4 months after last treatment).

Electro acupuncture: this part is very well detailed. Are there non-responders to this technique?
Response: As for any treatment including pharmacological treatment there are always non-responders. We estimate that around 75-80% respond to the treatment with experience from our pilot studies.

Primary outcome page 14 line 24: you are mixing 2 outcomes with two groups. PCOS vs Control at baseline and PCOS women randomized in the 3 groups with post vs pre evaluations after 4 months of the intervention. Please clarified your design. In this part, you could remove line 33 to 35.
Secondary outcome part: list of the outcome only with units. Remove all sentences as "all women with be examined by DXA to measure lean and fat mass and bone mineral density using a Lunar Prodigy Advance whole body scanner (GE Medical Systems)". This sentence have to be be add to "Study procedure" part.
Response: This has been done.

Sample size: based on HbA1c, my sample size calculation is different than 23. With a mean difference of 0.86 for HbA1c and a SD of 1.4 = 42 for each group.
Line 32 page 17: number of subject=23 subjects/groups as mentioned in "sample size" is not equal to 116.
Response: We agree that these calculations are wrong. After consulting a new statistical expert we realize that our sample size calculation is not entirely correct. We have therefore re-calculated and decided to have only one primary outcome variable = Hba1c, see page 14, Line 427-433. The main reason for this is that our statistical expert advise us not do an interim analyses as it is primarily to be used if you need to terminate the experiment due to detrimental effects.

Statistical analyses: "when the intended number of participants for Hba1c have been reached, an interim analyses will be performed".
Repeated Significance Tests with interim analysis according to Haybittle–Peto boundary is a rule for deciding when to stop a clinical trial prematurely. I appreciate this transparency statement by the authors. Interim analyses are statistically complex to implement but methodologically necessary.
Who is performing this interim analyses? Classical interim analyses are performed by an independent statistician who should be a person other than the regular study statistician. **Response:** As given above, we will not perform any interim analyses and this part has been deleted. See page 14-15.

Line 38 page 17: "The stop criterion are meant for both co-primary and covers two group comparisons". Witch two groups are you talking about? Acupuncture and Metformin group? Are you stopping the research if these two groups (Acup. vs Metf.) are equivalent in the HbA1c evolution according to one of your hypothesis? Or if one of these two groups (or both) will show a superiority vs Lifestyle group? This two hypothesis are different and not clear.
Line 43 page 17: May I have missed this information but "posthoc corrections as given below." are not mentioned below. Is it Bonferroni correction with 0.05/(number of groups) = 0.0167?

Response: No stop will be done. Correction has been added, page 15, line 459-461: As given above, Bonferroni corrections are very conservative and we have decided not to use it.

Trial Status part: the dates mentioned here are 2015, 2016 and 2019, with a gap between them. It could be informative to add the number of women already recruited in the middle of 2018 in each country.
Response: Number of participants randomized in Sweden: 26 and in China 48 in August 2018, see page 16 line 500-501.

How about women assessed for eligibility but not randomized because of declined to participate/pregnant … etc? A CONSORT flow diagram will be necessary for the main publication and should mentioned here.
Response: Yes, of course we keep track of all women and a CONSORT flow diagram will be included in the main publication. This is also given in page 16, line 488-489.

VERSION 2 – REVIEW

REVIEWER	Renato Pasquali UNIBO, Italy
REVIEW RETURNED	12-Sep-2018

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.