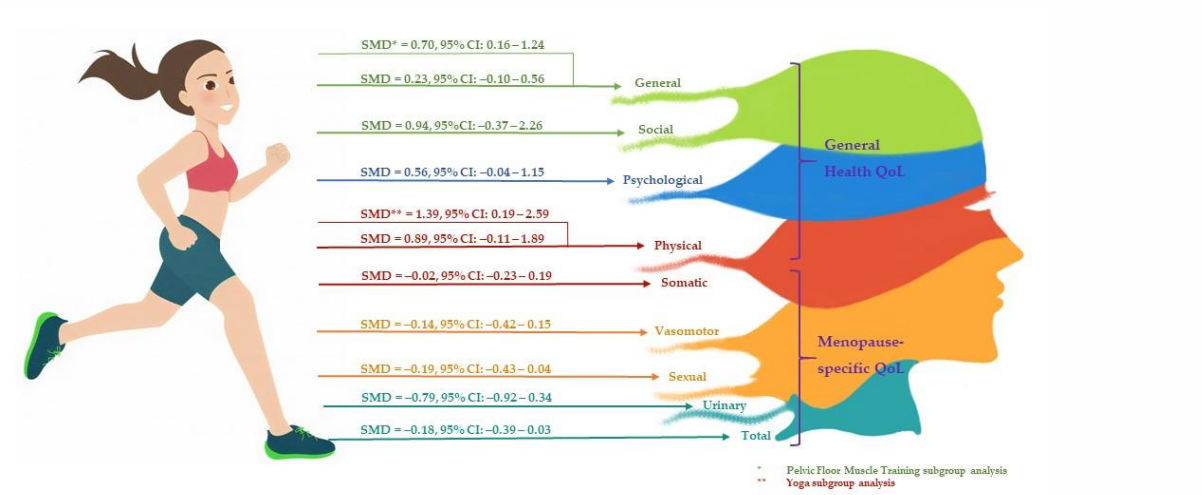


Response to Reviewer 1 Comments

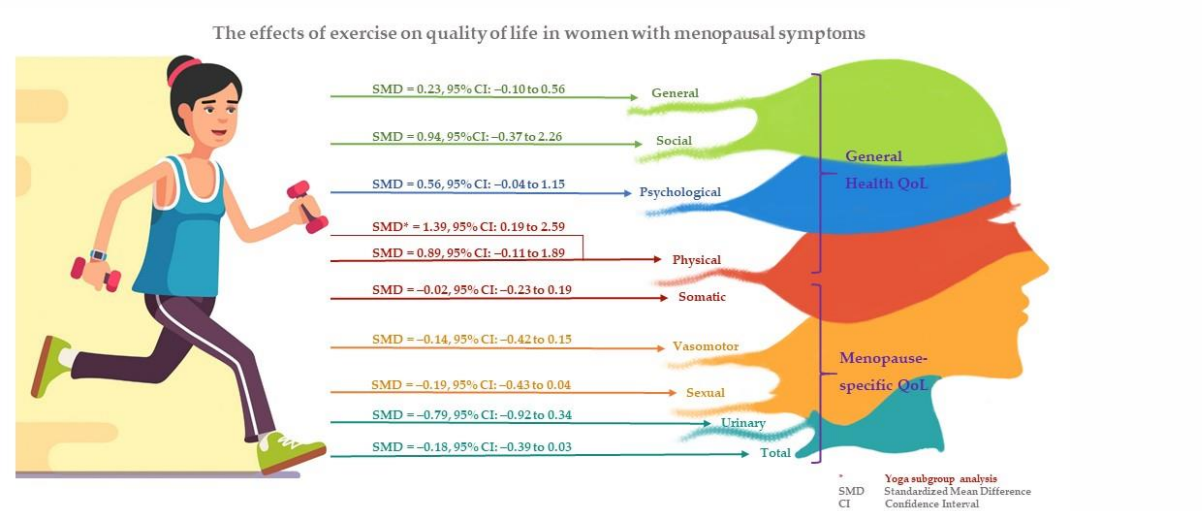
Point 1: Suggest moving figure to the results section, replace the runner with something more age-appropriate for women experiencing menopause.

Response 1: Thanks for your valuable comments. We replaced the runner with something more age-appropriate for women experiencing menopause. However, we did not move the figure to the results section because the results shown in the figure were also found in the results section in more detail. Because this figure is a graphical abstract summarizing and visualizing the main findings, we kept it in the abstract section after adding a short explanatory title to it (as suggested by another reviewer). In addition, after substituting a fixed-effect model with a random-effects model in subgroup analysis for the pelvic floor muscle training (as suggested by another reviewer), we updated the result in the graphical abstract. Please see the details below.

Before revision



After revision



Point 2: P.3 line 106, do you mean “conducted” instead of “conduced”.

Response 2: We changed ‘conduced’ into ‘conducted’.

Point 3: Line 126, missing a comma after “reporting bias”.

Response 3: Thank you for such a careful check. As suggested by another reviewer, we updated the risk-of-bias (ROB) assessment tool from the Cochrane ROB tool for RCTs (ROB 1) to the Cochrane revised ROB tool for RCTs (ROB 2). This led to a changed description as follows:

Before revision> Risk-of-bias was assessed independently by two researchers (T.M.N. and T.N.T.) using the Cochrane risk-of-bias tool for RCTs, which divides potential bias into six domains, selection bias, reporting bias attrition bias, detection bias, performance bias, and other bias. This tool also guided assessment of each domain as low, unclear, or high (Table S1).

After revision> Risk-of-bias (**ROB**) was assessed independently by two researchers (T.M.N. and T.N.T.) using the Cochrane **revised ROB** tool for RCTs, which divides potential bias into **five domains, ROBs produced from the randomization process, from deviations developed by the intended interventions, from missing outcome data, from different outcome measurement methods, and from preferable selection of the reported results.** This tool also guided assessment of **ROBs overall and by each domain as low ROB, some concerns, or high ROB.**

Point 4: Table 1, expand QoL domain headings.

Response 4: As you indicated, we expanded QoL domain headings in Table 1 as follows:

QoL questionnaire	Abbreviation	Type ²	QoL domain ³							
			General Health				Menopause-specific			
			General	Physical	Psychological	Social	Vasomotor	Sexual	Urinary	Somatic
Global QoL [37]	GQOL/VAS ¹	G	P	-	-	-	-	-	-	-
36-Item Short Form Health Survey [38]	SF-36	G	P	P	P	P	-	-	-	-
World Health Organization QoL – Brief Version [39]	WHOQOL-BREF	G	-	P	P	P	-	-	-	-
Utian QoL [40]	UQOL	S	P	-	N	-	-	-	-	-
Menopause-Specific QoL [41]	MENQOL	S	-	N	N	-	N	N	-	-
International Consultation on Incontinence Questionnaire (Short Form) [42]	ICIQ-SF	S	N	-	-	-	-	-	-	-
King’s Health Questionnaire [43]	KHQ	S	N	-	-	-	-	-	N	-
Sexual Activity Questionnaire [44]	SAQ	S	-	-	-	-	-	P	-	-
Greene Climacteric Scale [45]	GCS	S	-	-	-	-	N	N	-	N
Women’s Health Questionnaire [46]	WHQ	S	-	-	-	-	N	N	-	N
Hot Flash Related Daily Interference Scale [47]	HFRDIS	S	-	-	-	-	N	-	-	-
Hot Flash Rating Scale [48]	HFRS	S	-	-	-	-	N	-	-	-
Bristol Female Lower Urinary Tract Symptoms [49]	BFLUTS	S	-	-	-	-	-	-	N	-

Functional Assessment of Cancer Therapy for Endocrine Subscale [50]	FACT-ES	S	-	-	-	-	-	-	-	-	P
--	---------	---	---	---	---	---	---	---	---	---	---

QoL = Quality of Life;

¹ As being assessed via a single-item 100mm visual analogue scale (VAS), GQOL was also called VAS;.

² G = Generic QoL questionnaire; S = Menopause-specific QoL questionnaire;

³ P = Positive direction (Higher score denoted better QoL); N = Negative direction (Higher score denoted worse QoL).

Point 5: Lines 156-157 30% is repeated for I^2 .

Response 5: Thank you for pointing it out correctly. Because we found some errors in our description, we changed those correctly as follows:

Before revision> The magnitude of heterogeneity was classified as no heterogeneity ($I^2 = 0$ to 30%), moderate ($I^2 = 30$ to 49%), substantial ($I^2 = 50$ to 74%), or considerable ($I^2 = 75$ to 100%) [51].

After revision> The magnitude of heterogeneity was classified as **low** ($I^2 = 0$ to **24%**), moderate ($I^2 = 25$ to 49%), substantial ($I^2 = 50$ to 74%), or considerable ($I^2 = 75$ to 100%) heterogeneity [51].

Point 6: Table 2S for the number and why full texts were excluded is missing (from the editors).

Response 6: As we mentioned above, we changed risk-of-bias tool and thus deleted previous Table S1 because it explained previous risk-of-bias tool. Therefore, the previous Table S2 was renamed to Table S1 in the Supplementary Files submitted along with this revised manuscript. We also added numbers of the excluded full texts and explained the reason why some full texts were excluded in the text. The edited part was highlighted **in bold** as follows:

Before revision> Next, after excluding 81 publications such as non-English articles, conference abstracts, non-RCTs, and unavailable full texts, we included nine articles in the qualitative synthesis. Detailed reasons for full-text exclusion are presented in Table S2.

After revision> **After screening 90 possibly relevant studies as full texts, we further excluded 81 publications including (1) eight non-RCTs; (2) twelve unavailable full texts; (3) four conference abstracts; (4) six non-English articles; (5) five studies combining exercise with other interventions; (6) one study in which both the intervention and the control groups performed exercises; (7) one study using active treatment control group; (8) three studies not providing QoL quantitative outcomes; (9) six studies of menopausal women that did not mention any menopausal symptoms as inclusion criteria; (10) five studies published as different perspectives of the same project with the included studies; (11) 30 studies having a combination of the aforementioned reasons.** Detailed reasons for full-text exclusion are presented in Table S1.

Point 7: Line 192, “gymnastics”-this is not something that menopausal women typically do, please explain.

Response 7: The reviewer was right that gymnastics is not a commonly-used type of exercise among menopausal women. However, it is undeniable that choosing which type of exercise to do is a personal choice. Menopausal women who are interested in gymnastics are likely to do so. In fact, there are many types of gymnastics which are modified to meet the physical conditions of middle-aged or aged adults such as silver gymnastic exercise and tai-chi gymnastics.

However, the reason why we mentioned gymnastics is that it is a non-mandatory part of the unsupervised aerobic training program used in the study by Luoto et al. [30], whereas the aim of section 3.2 is to briefly describe the characteristics of the included studies. To clarify this idea, we added the reference sign [30] at the end of the sentence. For further details, we would like to quote the description of the intervention from the original study as follows:

“The unsupervised aerobic training programme included aerobic training four times per week, with 50 minutes of exercise each time... At least two sessions had to include walking or Nordic walking, and the other two sessions could include walking, Nordic walking, jogging, cycling, swimming, skiing, aerobics, or other gymnastic exercise”.

Points 8–9:

Point 8: Table 2, providing minutes of exercise would be better than number of bouts.

Point 9: Table 2, results are inconsistently reported.

Responses 8–9: Thank you for pointing this out. We are not sure that we understood this comment correctly. If the reviewer mentioned the column for frequency, we would like to explain below. If not, please point out the inconsistency in details, we will check it out again.

First, we added the exercise frequency for the Ngowsiri et al.’s study, which is 90 minutes per bout and 3 bouts per week. Regarding other included studies, to help readers to easily understand the frequency of exercise, we provided both the minutes and the number of bouts. As the frequency of exercise was described differently among the nine included studies, to increase the consistency, we converted all hour data to minute, month to week, and day to bout if relevant. In case these data are not sufficiently mentioned in the included studies (i.e. home-based interventions are often more flexible), we tried to present them with the highest consistency. In addition, as suggested by other reviewers, we added the dropout rate, the number of participants in each group, and the full names of all the abbreviations used in the table as well as deleted all the years to meet the numbering style of the journal. Please see all changes in Table 2 highlighted in red below.

Author, Year	Participant	Country	Dropout rate	Intervention	Frequency ¹	Duration ¹	Control	QoL questionnaire	Result
Mercier et al. [24]	32 women with GSM; ≥ 55 years old	Canada	9.4% (3/32)	PFMT; Supervised and home-based (n = 32)	Supervised: 1b/w, 60min/b Home-based: 5d/w	12w	None (n = 0)	ICIQ-VS	QoL and sexual function of women with GSM improved after the intervention.
Bertotto et al. [25]	49 postmenopausal women with urinary symptoms; 50–65 years old	Brazil	8.2% (4/49)	PFMT: contraction (n = 15)	2b/w, 20min/b	4w	No treatment (n = 14)	ICIQ-SF	The PFMT group exhibited significant increases in ICIQ-SF scores.
Ngowsiri et al. [26]	54 menopausal women with menopausal symptoms; 45–59 years old	Thailand	7.4% (4/54)	Rusie Dutton dance of 16 yoga-like postures; Supervised (n = 24)	3b/w, 90min/b	13w	Provided a handbook (n = 26)	MENQOL	There was a significant improvement in all MENQOL domains in the experiment group and between the two groups.
Jayabharathi et al. [27]	260 women with menopausal symptoms; 45–55 years old	India	2.3% (6/260)	Yoga; Supervised and home-based (n = 128)	Supervised: 5 consecutive days: 2b/d, ~45min/b; Later: 2d/w Home-based: 35-40min/d	18w	No intervention (n = 126)	WHOQOL-BREF	A statistically significant difference between the study group and the control group was observed in terms of all domains of QoL.
Avis et al. [28]	54 menopausal women with ≥ 4 hot flashes/day; 45–58 years old	USA	20.4% (11/54)	Integral Yoga Supervised and home-based qualified DVD for self-practice (n = 15)	Supervised: 1b/w, 90min/b Home-based: 3b/w, 15min/b	10w	Waitlist (n = 14)	SF-36 HFRDIS VAS	Yoga can act as a behavioral option which helps reducing hot flashes. There was no advantage of Yoga over other types of exercise.
Pereira et al. [29]	45 postmenopausal women with urinary symptoms; 53–73 years old	Brazil	8.9% (4/45)	PFMT: contraction; Supervised by a physical therapist (n = 15)	2b/w, 40min/b	6w	No intervention (n = 15)	KHQ	Several positive results of PFMT in treatment for urinary leakage, pelvic floor muscle pressure, and QoL were observed.
Luoto et al. [30]	176 menopausal women with daily hot flashes;	Finland	12.5%	Unsupervised aerobic training	4b/w, 50min/b	24w	1-2 lectures/month;	SF-36 WHQ	Women in the intervention group had significantly

	40–63 years old		(22/176)	(n = 74)			60-75min (n = 80)		higher SF36 scores in mental health than those in the control group. There were significant differences in improvement for menopause symptoms, SAQ and SF-36 between the intervention and the control group.
Duijts et al. [31]	422 breast cancer patients with menopause symptoms; 48.2 ± 5.6 years old	The Netherlands	16.6% (70/422)	Home-based, self-directed exercise program; Assisted by a physiotherapist (n = 104)	Home-based: 150-180min/w	12w	Waitlist (n = 103)	SF-36 HFRS SAQ FACT-ES BFLUTS	
Elavsky et al. [32]	164 sedentary women with menopausal symptoms; 42–58 years old	USA	24.4% (40/164)	(1) Walking: supervised; (n = 63) (2) Iyengar Yoga: supervised (n = 62)	(1) 3b/w; 60min/b; (2) 2b/w; 90min/b	16w	Waitlist (n = 39)	UQOL GCS	The yoga and walking interventions showed positive effects on menopause-specific QoL.

GSM = Genitourinary Syndrome of Menopause; PFMT = Pelvic Floor Muscle Training; QoL = Quality of Life; n = number of participants;.

ICIQ-VS = International Consultation Incontinence Questionnaire – Vaginal Symptoms; ICIQ-SF = International Consultation on Incontinence Questionnaire (Short Form); MENQOL = Menopause-Specific Quality of Life; WHOQOL-BREF = World Health Organization QoL – Brief Version; SF-36 = 36-Item Short Form Health Survey; HFRDIS = Hot Flash Related Daily Interference Scale; VAS = Visual Analogue Scale; KHQ = King's Health Questionnaire; WHQ = Women's Health Questionnaire; HFRS = Hot Flash Rating Scale; SAQ = Sexual Activity Questionnaire; FACT-ES = Functional Assessment of Cancer Therapy for Endocrine Subscale; BFLUTS = Bristol Female Lower Urinary Tract Symptoms; UQOL = Utian Quality of Life; GCS = Greene Climacteric Scale;.

¹ b = bout; d = day; w = week; min = minute.

Point 10: Start of section 3.5, aren't the results statistically significant $p=0.01$?

Response 10: Although a fixed-effect model showed a statistically significant association between PFMT and general QoL (SMD = 0.70, 95% CI: 0.16 to 1.24), its usage seems debatable because of an unexplained considerable heterogeneity ($I^2 = 78\%$). Besides, although two studies [25,29] included in such analysis both recruited Brazilian women with urinary symptoms, used two supervised PFMTs per week as the intervention compared with no treatment, and used a similar question to measure general QoL, the true intervention effects were rarely similar in studies conducted independently. Taken these points into account, as suggested by another reviewer, we decided to use a random-effects model to analyse the effects of PFMT on QoL in women with urinary symptoms. This led to a different SMD, 95% CI, and p -value. Therefore, we edited the text in the Subgroup analyses section. Please find the updated result as well as the interpretation highlighted **in bold** below.

Before revision> There was a significant positive improvement in general QoL in women with urinary symptoms after PFMT interventions compared with no intervention (SMD = 0.7, 95% CI: 0.16 to 1.24, $p = 0.01$).

Table 3. Meta-analyses of the effects of PFMT and yoga on QoL domains.

QoL domain	No. studies	No. participants (exercise)	No. participants (control)	Model	SMD (95% CI)	p -Value	Heterogeneity I^2 ; χ^2 ; p -Value
PFMT							
General	2	30	29	F	0.70 (0.16 to 1.24)	0.01	78%; 4.52; 0.03
Yoga							
General	2	76	53	R	-0.07 (-0.46 to 0.33)	0.74	12%; 1.14; 0.29
Physical	3	167	166	R	1.39 (0.19 to 2.59)	0.02	93%; 29.31; p^*
Psychological	4	228	205	R	0.76 (-0.3 to 1.81)	0.16	95%; 63.01; p^*
Sexual	2	85	65	R	-0.36 (-1.18 to 0.46)	0.39	81%; 5.39; 0.02
Vasomotor	3	100	79	R	-0.37 (-1.15 to 0.4)	0.34	82%; 11.18; p^{**}

F = Fix-effect model; R = Random-effects model; $*p < 0.00001$; $**p < 0.01$.

After revision> **We found a non-significant** improvement in general QoL in women with urinary symptoms after PFMT interventions compared with no intervention (SMD = **0.76**, 95% CI: **-0.40** to **1.92**, $p = 0.20$).

Table 3. Meta-analyses of the effects of PFMT and yoga on QoL domains.

QoL domain	No. studies	No. participants (exercise)	No. participants (control)	SMD (95% CI)	p -Value	Heterogeneity I^2 ; χ^2 ; p -Value
PFMT						
General	2	30	29	0.76 (-0.40 to 1.92)	0.20	78%; 4.52; 0.03
Yoga						
General	2	76	53	-0.07 (-0.46 to 0.33)	0.74	12%; 1.14; 0.29
Physical	3	167	166	1.39 (0.19 to 2.59)	0.02	93%; 29.31; p^*

Psychological	4	228	205	0.76 (-0.3 to 1.81)	0.16	95%; 63.01; p^*
Sexual	2	85	65	-0.36 (-1.18 to 0.46)	0.39	81%; 5.39; 0.02
Vasomotor	3	100	79	-0.37 (-1.15 to 0.4)	0.34	82%; 11.18; p^{**}

QoL = Quality of Life; PFMT = Pelvic Floor Muscle Training;

SMD = Standardized mean difference; CI = Confidence Interval;

¹ $p < 0.00001$; $p^{**} < 0.01$.

Point 11: Table 4, there is no explanation for “other bias” despite the fact that the category was unanimously “high”.

Response 11: Thank you for pointing this out. Initially, the other biases include the lack of a control group [24], the limited sample size [24,25,28,29], brief follow-up period [25], the possibility of some degree of placebo effect [26,31], using an invalidated instrument for baseline menopausal symptoms assessment [30], and the effects of uncontrollable healthy lifestyle practices [27], as well as challenges in the recruitment, the intervention attendance, and the follow-up for the home-based practices [28].

However, as mentioned in **Response 3**, we updated the ROB assessment tool. Compared with the ROB 1 which does not describe the specific issues addressed and thus causes confusion, the ROB 2 not only covers all types of bias that can affect results of randomized trials but also provides algorithms that map responses to signalling questions onto domain-level judgements as well as overall ROB judgements. The signalling questions covered all aforementioned other biases in the ROB 1. Updating the ROB tool led to changes in ROB assessments and in interpretations of the results.

In addition, as suggested by another reviewer, we deleted all the years in Table 4 to meet the numbering style of the journal. Edited text and table were as follows;

Before revision> Table 4 summarizes the risk-of-bias assessment. In general, the risk of selection bias of the included studies was low. Six studies [24,25,27,29–31] carefully described randomization, while five articles [25,29–32] reported adequate allocation concealment. The majority of studies reported an unclear or high risk of detection and performance biases, which were referred to as the ability to blind participants and personnel as well as outcome assessors. In contrast, attrition and reporting biases were low in seven [24,27–32] and eight [24–30,32], respectively, of the nine selected studies. Only Duijts et al. (2012) [31] reported a high risk of selective reporting.

Table 4. Risk-of-bias assessment of the included studies.

Author, Year	Selection bias ¹	Selection bias ²	Reporting bias ³	Attrition bias ⁴	Detection bias ⁵	Performance bias ⁶	Other bias
Mercier et al., 2019 [24]	Low	Unclear	Low	Low	Unclear	High	High
Bertotto et al., 2017 [25]	Low	Low	Low	Unclear	High	High	High
Ngowsiri et al., 2014 [26]	Unclear	Unclear	Low	Unclear	Unclear	Unclear	High
Jayabharathi et al., 2014 [27]	Low	Unclear	Low	Low	Unclear	High	High
Avis et al., 2014 [28]	Unclear	Unclear	Low	Low	Unclear	Unclear	High
Pereira et al., 2012 [29]	Low	Low	Low	Low	High	Low	High
Luoto et al., 2012 [30]	Low	Low	Low	Low	Unclear	Unclear	High
Duijts et al., 2012 [31]	Low	Low	High	Low	Unclear	High	High
Elavsky et al., 2007 [32]	Unclear	Low	Low	Low	Unclear	Low	High

¹ Random sequence generation; ² Allocation concealment; ³ Selective reporting; ⁴ Incomplete outcome data; ⁵ Blinding outcome assessment; ⁶ Blinding participants & personnel.

After revision> Table 4 summarizes the ROB assessment. Except the study by Luoto et al. [30], judgements about the overall ROB of eight remaining RCTs express some concerns. The ROB_s relating to selection of the reported result and randomization process were low in all and in most of the included studies, respectively. Despite mentioning a random assignment of participants to exercise and control groups, two studies [26,27] insufficiently described the process. Despite relatively high dropout rates (Table 2), five [25,26,28,30,32] out of nine studies did not provide adequate evidences that the results were not biased by missing outcome data, reporting some concerns about this ROB domain. In general, the ROB_s due to deviations from intended interventions and measurement of the outcome were considerable.

Table 4. Risk-of-bias assessment of the included studies.

Author, Year	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result	Overall
Mercier et al. [24]	Low	Some concerns	Low	Low	Low	Some concerns
Bertotto et al. [25]	Low	Some concerns	Some concerns	Some concerns	Low	Some concerns
Ngowsiri et al. [26]	Some concerns	Some concerns	Some concerns	Some concerns	Low	Some concerns
Jayabharathi et al. [27]	Low	Some concerns	Low	Some concerns	Low	Some concerns
Avis et al. [28]	Some concerns	Some concerns	Some concerns	Some concerns	Low	Some concerns
Pereira et al. [29]	Low	Some concerns	Low	Some concerns	Low	Some concerns
Luoto et al. [30]	Low	High	Some concerns	Some concerns	Low	High
Duijts et al. [31]	Low	Some concerns	Low	Some concerns	Low	Some concerns
Elavsky et al. [32]	Low	Some concerns	Some concerns	Low	Low	Some concerns

Point 12: Conclusion should recommend exercise that is specific to menopause, i.e. B.K.S. Iyengar has yoga practices specifically scripted for menopausal women. Consult the text “B.K.S. Iyengar Yoga: The Path to Holistic Health”

Response 12: Thank you for the suggestion. Based on the previous findings, we suggest that **supervised** PFMT should be prescribed as a conservative therapy for women with menopausal symptoms.

However, because of the changes mentioned in **Response 10**, there were not enough evidence for the effects of supervised PFMT and yoga on QoL in women with menopausal symptoms. This suggests that well-designed trials are needed to confirm the effects of exercise in general. Along with the updated findings, we edited the conclusions as follows;

Before revision> This systematic review found marginally significant improvements in physical and psychological QoL scores, but no significant improvements in general, social, and menopause-specific QoL scores, in women with menopausal symptoms after exercise interventions compared with no active interventions. Yoga and PFMT were the most common interventions used in the included studies. Based on a significant positive association between PFMT and general QoL in women with urinary symptoms, we suggest that PFMT should be prescribed as a conservative therapy for such women. Because the effects of yoga on QoL in women with menopausal symptoms are inconsistent, more well-designed studies that investigate similar types of yoga are needed.

After revision> This systematic review found **some** improvements in physical and psychological QoL scores, but no significant improvements in general, social, and menopause-specific QoL scores, in women with menopausal symptoms after exercise interventions compared with no active interventions. Yoga and PFMT were **respectively** the most common interventions used **for women with menopausal and urinary symptoms** in the included studies. **Although yoga significantly improved physical QoL, its effects on general, psychological, sexual, and vasomotor symptoms QoL were inconclusive. PFMT non-significantly improved general QoL. Our findings suggest that well-designed trials are needed to confirm the effects of exercise on QoL in women with menopausal symptoms.**