PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Integration of Sex and Gender in a Continuing Professional Development Course on Diabetes and Depression: A Mixed Methods Feasibility Study
AUTHORS	Deom Tardif, Alèxe; Gogovor, Amédé; Guay-Bélanger, Sabrina; Audet, Denis; Parent, Nicole; Gaudreau, André; Remy-Lamarche, Danièle; Vigneault, Luc; Ngueta, Gérard; Bilodeau, André; Légaré, France; mATrICES-F Group, the

VERSION 1 – REVIEW

REVIEWER	Alva, Maria Georgetown University, Massive Data Institute
REVIEW RETURNED	07-Jun-2021

GENERAL COMMENTS	• The authors take the premise that gender matters as given. For a general audience, this should have been better explained. For example, there is work showing that demographic characteristics matter very little in determining the prevalence of diabetes and depression and often are correlated in the opposite direction. What matters significantly more are SES indicators. The evidence to justify sex and gender integration in clinical practice needs to be spelled out in greater detail. What evidence is there that males and females react differently to diabetes or deal differently with depression?
	  • The statement “Twice as many women suffer from depression” -- should this read instead as “twice as many women are diagnosed with depression or seek help compared to men”? Especially a context where the authors cite that “three times as many men commit suicide.”
	  • A needs assessment of physicians is mentioned on page 9, line 45. What were the findings, and how they motivate this intervention? If previously published, please provide a reference.
	  • What is the epidemiological data on the differences in incidence, prevalence, morbidity, and mortality between men and women diagnosed with both T2D and depression? This information should be available since it is part of the training health professionals receive.
	• The setting with a treatment and a control group is that of an RCT but the authors stressed that it is not. Why was an RCT not feasible?

	• Could the authors offer references for the theory of planned behavior, theoretical domains framework, and the COM-B model?
	  • On page 8, spell out COM-B
	• Given that the invitation was done via email and registration forms, it seems possible to have a well-defined starting sample -- please illustrate if that is not the case. Because one of the main outcomes was recruitment, it would have been useful to know how many people participated in the study out of those who were emailed about participating and not just those who took the training out of the ones who participated (or is that all took the training but only a fraction participated in the study)?
	  • The intervention needs to be better defined. Is the training the 1h and 30 mins and the participation is the answering of questionnaires?
	  • On page 16, the authors report 12 courses (is each course 1h and 30 mins?), where the provinces expanded or the number of sessions within the provinces expanded?
	• Page 14, line 42 onwards. Why were not all questionnaires analyzed? And why were different proportions analyzed in the treatment and control groups?
	• Why not using the age of the participant as a continuous variable?
	• The point of table 1 should be explained -- if there is no balance between treatment and control, what is the point of the control group?
	• Explain what it means a -0.57 decrease in intention. Would it not be better to test for changes that move the response in the Likert scale by one echelon?
	• Table 2 needs more interpretation. How and why does the intervention change psychological determinants?
	• Table 4. It would be helpful to have more explanation on how the TDF domains link to the intervention. I also struggled to see how skills linked to any of the illustrative quotes of the facilitator and barrier explanation.
	• Sex and gender are routinely recorded and clinical practice, but they do not seem to be considered when administering care unless the patient is undergoing menopause. It is unclear how the authors are advocating this intervention enters clinical consideration. How can this intervention speak to other countries/areas outside Quebec?

REVIEWER	Ahmed, Sofia University of Calgary
REVIEW RETURNED	09-Sep-2021

GENERAL COMMENTS	The manuscript by Dr. Déome Tardif and colleagues sought to demonstrate feasibility of integration of sex and gender (SG) in a continuing professional development course on diabetes and depression using a mixed methods approach. This non-randomized two-armed study across 3 Canadian provinces included health professionals who were assigned to either a CPD course with SG considerations or a similar course without these considerations. The main outcomes of interest feasibility of recruitment and intention of CPD organisations and

	patient partners, adherence to planned activities, health professionals' intention to include SG considerations in practice as measured by CPD reaction questionnaire, and barriers and facilitators using the Theoretical Domains Framework. The authors concluded that the intervention was feasible and may prompt health providers to modify their care delivery. The authors also identified facilitators and barriers using semi-structured group discussions. This is an interesting and important area of research. This work highlights the feasibility of integrating sex and gender into continuing professional development for health care professionals with the potential to translate into clinical practice and improve health outcomes. My suggestions to strengthen the manuscript are as follows: 1) The goal of the study was to show feasibility, which the authors have demonstrated through recruitment and fidelity of the intervention. The conclusions of the study state that the intervention "had a favourable impact on health care professionals' intention to include sex and gender considerations in caring for individuals with T2D and depression". However, have the authors considered how the non-randomized nature of the study which included participant preferences as well as the lack of pre-intervention measures of intent could influence the findings? 2) There are multiple subgroup analyses and comparisons, which introduces the potential for a false positive. The authors state the effectiveness of the intervention was greatest among older and rural-based physicians (page 23) but have the authors considered that the study was not designed or powered to show this definitively? 3) The terms and definitions sex and gender appear to be used interchangeably at times in the manuscript. For example, page 8 "two males and one female" patient partners, page 10, line 46 "efforts were made to equally divide groups regrading number and sex of participants", line 53 "same two physicians (one male, assigned to the control group, and one female". Do the authors mean to identify the sex (rather than the gender identity)? Table 1 – should it be gender identity instead of sex? Similarly, page 23 line 18 "gender considerations" but then discuss female sexual health and cardiovascular risk, which are sex-based considerations. 4) Have the authors considered that the participants' preference was taken into account in assignation to a group may have influenced the results due to bias? Similarly, could the gender identity of the group leader have influenced the participants' stated intentions? 5) Are there studies that show stated intentions as measured by the CPD reaction questionnaire translate into clinical practice changes?

VERSION 1 – AUTHOR RESPONSE

Reviewers' comments	Response to reviewer	Location of change

Reviewer 1		
1. The authors take the premise that gender matters as given. For a general audience, this should have been better explained. For example, there is work showing that demographic characteristics matter very little in determining the prevalence of diabetes and depression and often are correlated in the opposite direction. What matters significantly more are SES indicators. The evidence to justify sex and gender integration in clinical practice needs to be spelled out in greater detail. **What evidence is there that males and females react differently to diabetes or deal differently with depression?**	Thank you for this comment. We have spelled out the rationale for sex and gender integration in clinical practice (and consequently, in CPD) in greater detail in the Introduction and updated the evidence, as follows: "*During their lifetime, women are twice as likely as men to be diagnosed with depression. In contrast, three times as many men commit suicide (5, 10, 11). Recent evidence supports a link between type 2 diabetes (T2D) and depression, and shows that sex and gender are influential factors in this comorbidity (7, 9). The prevalence of depression in diabetic patients is higher in females than males (23.8% and 12.8%, respectively) (7). On the other hand, a pooled result from 32 studies described that the risk of developing T2D in patients diagnosed with depression is higher in men than in women (RC=1.63 vs RC=1.29, respectively) (7, 12, 13). The differences are explained by biological differences and psychosocial factors such as body mass index, differences in the distribution of types of adipose tissue, an imbalance of sex hormones, socioeconomic status, psychosocial stress and sleep deprivation (7, 9). Co-morbidity and mortality associated with the complications of T2D and depression are also different for men and women. For instance, men develop diabetic food syndrome at earlier ages and are more likely to have complications leading to amputations (7, 14). Women, on the other hand, have a higher risk of metabolic syndrome and fatal coronary heart disease than men (7, 15, 16). T2D and depression are also affected by gender differences. This gap could be explained in part by the different behaviours associated gender representations of men and women, as well as their different perceptions of stress (17-19).* We also highlighted how integration of sex and gender considerations in CPD activities is a promising avenue for reducing inequities between men and women: "*Despite the impacts of sex and gender differences on prevalence, diagnosis, treatment, outcomes, and equity, evidence on the importance of these differences has yet to be translated adequately into clinical training or practice (2, 5, 20).*	p.7

	… *We argue that integrating sex and gender considerations into CPD is a promising avenue for addressing the inequities between men and women (5).* In the Intervention section we added this sentence with references: *There is growing evidence of a link between T2D and depression and the importance of sex as a risk factor for this comorbidity (34-36).*	p. 8 p. 8 p. 10
1.1. The statement "Twice as many women suffer from depression" -- should this read instead as "twice as many women are diagnosed with depression or seek help compared to men"? Especially a context where the authors cite that "three times as many men commit suicide."	Thank you for pointing this out. Indeed, the contrast is startling and has to do with gender differences in dealing with depression. We have corrected the statement to read: *"During their lifetime women are twice as likely as men to be diagnosed with depression. In contrast, three times as many men commit suicide (5, 10, 11)."*	p. 7
1.2. A needs assessment of physicians is mentioned on page 9, line 45. What were the findings, and how they motivate this intervention? If previously published, please provide a reference.	Thank you for this question. The needs assessment was done with CPD stakeholders engaged in the project. It has not been published as these are internal documents for this organisation. We clarified and added details as follows: "*Informed by a continuing medical education needs assessment by our key CPD stakeholder and partner*, Médecins francophones du Canada *(data not published),* we chose patients with T2D and depression combined as the clinical topic, *as physicians felt there was a gap in their education about this comorbidity.*"	p. 10

1.3. What is the epidemiological data on the differences in incidence, prevalence, morbidity, and mortality between men and women diagnosed with both T2D and depression? This information should be available since it is part of the training health professionals receive.	See our answer to your first question, above.	p. 7
2. The setting with a treatment and a control group is that of an RCT but the authors stressed that it is not. Why was an RCT not feasible?	Most of our courses were given in family medicine groups or hospitals and embedded in existing CPD/CME scheduled events, where participants were not asked to register in advance and just showed up on the day of the activity. For courses embedded in a CME conference that were publicised online beforehand, we had to respect Médecins francophones du Canada's conference guidelines and allow participants to choose their courses (both were advertised similarly). However, we made sure participants were blinded to the intervention and control group once they got to the door of the event; objectives presented in the online program were the same, and apart from the content itself, only the teacher differed between courses. We also accepted non-registered participants the day of the activity so as to offer the courses to as many participants as possible. At the end, even without randomization, interestingly both our intervention and control groups ended up well balanced regarding sociodemographic data (see Table 1). We added this sentence to the limitations of the study: *"Second, the fact that participants could choose which course to attend (according to conference guidelines), and hence the non-randomized nature of the study, may have biased our feasibility findings."*	p. 24
3. Could the authors offer references for the theory of planned behavior, theoretical domains framework, and the COM-B model? • On page 8, spell out COM-B	Thank you, we have added references for our theoretical frameworks and spelled out COM-B as follows: "*We used the Theory of Planned Behavior for quantitative analysis (27, 28), the Theoretical Domains Framework (TDF) for qualitative analysis (29, 30), and*	p. 9

	the COM-B (Capability, Opportunity, Motivation and Behavior) model to triangulate findings (31)."	
4. Given that the invitation was done via email and registration forms, it seems possible to have a well-defined starting sample -- please illustrate if that is not the case. Because one of the main outcomes was recruitment, it would have been useful to know how many people participated in the study out of those who were emailed about participating and not just those who took the training out of the ones who participated (or is that all took the training but only a fraction participated in the study)?	Thank you for this comment. Participants were those who registered in the CPD course, signed consent, completed the course, filled out the evaluation questionnaire and, where the discussion was offered, took part in the 30-minute discussion afterwards. All took the training but not everyone filled out a consent form – hence the data couldn't be added to our study. We clarified Figure 1 to better represent how many people were invited to participate in the study, how many received the training, and how many participated in the study (i.e. signed consent forms and completed the questionnaire). As now indicated in Figure 1: *Potential eligible healthcare professionals invited to participate in the study* (n=3344) And regarding 2000 employees of the Montfort Hospital: * *This is an approximate figure given* the changing dynamics of the hospital's professional environment; an email was sent to 2000 healthcare professionals, *others were invited using posters in the training sites, oral communication at a meeting with the organizing team of the clinical setting, and announcements in Médecins francophones du Canada's conference calendar.* As illustrated in Figure 1, 178 healthcare professionals received the training (intervention = 92, control = 86). At the end, 127 healthcare professionals participated in the study (accepted to sign consent forms and completed questionnaires). We therefore analyzed data from 127 participants (intervention = 49 and control = 78).	See file "Figure 1"
4.1. The intervention needs to be better defined. Is the training the 1h and 30 mins and the participation is the answering of questionnaires?	The intervention included the course that included the sex and gender content, the questionnaire and the discussion. We clarified on page 11 as follows: "Each course *(both control and intervention)* was a 45-minute lecture on T2D and depression followed by 15 minutes to fill in the CPD-Reaction questionnaire. An additional 30 minutes was planned for the semi-structured group discussion."	p. 11

4.2. On page 16, the authors report 12 courses (is each course 1h and 30 mins?), where the provinces expanded or the number of sessions within the provinces expanded?	The timing of each of the 12 courses was as above. We have clarified the number of sessions as follows: "Our CPD trainings were held in the three provinces as planned. We gave 12 courses instead of the six initially planned, *as additional organizations in Quebec City (n=1) and Montreal (n=2)* showed interest*. Due to time constraints imposed by CME settings, completing 1.5 hours (45-min course, 15-min evaluation and 30-min discussion) in all settings was not possible, therefore we held the group discussions in only two out of the six settings (Montreal and Ottawa).*	p. 15
5. Page 14, line 42 onwards. Why were not all questionnaires analyzed? And why were different proportions analyzed in the treatment and control groups?	Due to ethics guidelines, we only analyzed questionnaires from participants who agreed to sign a consent form. Since 127 participants signed the consent form, we included 127 questionnaires in our analyses. In cases where questionnaires were filled out but the consent form was not signed, we did not include these participants in our analyses. We added the following sentence to the Discussion: *"Also, due to ethics guidelines, we only analyzed questionnaires completed by participants who had also signed consent forms."* The different proportions analyzed in the treatment and control groups is explained by the fact that more participants agreed to sign the consent form in the control group than in the intervention group. This limitation is explained in the Discussion as follows: "Although the human resources for both groups were the same (trainer, research-assistant and patient-partners), the control group had an extra team member, resulting in unequal numbers of participants who signed consent in each group. The presence of this extra member could also explain the difference in the number of questionnaires collected in the two groups."	p. 24 p. 25

6. Why not using the age of the participant as a continuous variable?	We thank you for this comment and agree that it would have been better to use the age of participants as a continuous variable. However, we had to comply with ethics committees of three provinces (4 cities), each with their own rules. We were not allowed to directly ask participants their age, and therefore had to use ranges. Then, when we combined the data for the three provinces, the age ranges were different and the only age ranges we were able to use were < 44 and ≥ 45. In itself, this also reflects some of the particularities of doing research in a real applied CPD/CME environment (pragmatic trial).	
7. The point of table 1 should be explained -- if there is no balance between treatment and control, what is the point of the control group?	We explained above how it happened that there were different numbers in the group. We agree that this is an important limitation. However, we performed sensitivity analyses with all questionnaires we received (total n=159, intervention = 75 and control = 84). These analyses showed similar results to those presented in the manuscript. Moreover, sociodemographic data presented in Table 1 are balanced in each group. Finally, since we only had a post-intervention measure, the control group was necessary to measure the impact of the intervention on health professionals' intention to include sex and gender considerations in patient care.	
8. Explain what it means a -0.57 decrease in intention. Would it not be better to test for changes that move the response in the Likert scale by one echelon?	We are grateful to the reviewer for this relevant comment. It would be interesting to test for changes that move the response in the Likert scale by one echelon in the context of pre-intervention measures. In our study, in the absence of pre-intervention measures, we were obliged to use a control group to measure the impact of the intervention. To do this, it was necessary to calculate a mean difference in intention scores between the intervention and control groups. The value -0.57 corresponds to this difference. Please note that this is the between-group difference in the intention score, as observed post-intervention. It is not a "decrease" since we used a single post-intervention measure. In post-intervention, the mean score in the intervention group was -0.57 lower than the mean score in the control group (95%CI: -1.09 to -0.05; P=0.03). Given the lack of pre-intervention measures, we could not test for changes.	
9. Table 2 needs more interpretation. How and why does the intervention change psychological determinants?	Implementing and advocating for new practices is challenging and requires changes in the behavior of health professionals. Understanding the determinants of current and desired behaviors is therefore necessary. Both the CPD-Reaction questionnaire and the Theoretical Domains Framework (TDF) were developed by a collaboration of behavioral scientists	

	and implementation researchers who identified various theories relevant to implementation and grouped constructs from these theories into psychosocial determinants of behavior and domains (some of which converge in CPD-Reaction and the TDF). We expanded our explanation and gave an example to better explain results presented in Table 2 as follows: *"The intervention aims to influence behaviour by modifying intention and its psychosocial determinants. For example, the intervention could change beliefs about capabilities (or confidence), by increasing health professionals' knowledge about the desired behavior.* Table 2 shows scores for intention and its psychosocial determinants *for intervention and control groups* as evaluated using the CPD-Reaction questionnaire." We also added the following in the Discussion: "*Third, beliefs about capabilities as a facilitator showed the strongest mean difference between the intervention and control groups. These results are consistent with a literature review of 277 studies showing that the mechanisms of action most frequently associated with behaviour change techniques are beliefs about capabilities and intention (58). Adding a practical component to the CPD course could strengthen beliefs about capabilities."*	p. 16 p. 23
10. Table 4. It would be helpful to have more explanation on how the TDF domains link to the intervention. I also struggled to see how skills linked to any of the illustrative quotes of the facilitator and barrier explanation.	We used the following definition of Skills provided by the TDF to link to our qualitative code: **Skills (An ability or proficiency acquired through practice):** Skills development Competence Ability	

	Interpersonal skills Practice Skill assessment To clarify how we used the TDF domains, we rewrote the section on barriers and facilitators and our coding regarding knowledge and skills in the Results (Qualitative Findings) as follows: "***Barriers and facilitators mapped to the TDF domains*** *Ten barriers mapped to nine of the 14 TDF domains and seven facilitators mapped onto six of the domains. The most frequent barriers were related to Skills (e.g. failing to consider a patient's gender) (n=3) and to Social Influence (e.g. making gender assumptions about employment) (n=3). The most frequent facilitators were also related to Skills (n=4) (Table 4).* *We mapped to the Skills domain when the participants asked whether their patient was a woman or man before analyzing the clinical vignette, or else failed to ask the question (the fictive name of the patient – Dominique – was strategically ambiguous). Thus, failure to ask was coded as a barrier, and asking was coded as a facilitator. Discussion about information on sex and/or gender was coded as a facilitator in the Knowledge domain, but reporting differentiating between women and men patients in clinical practice was coded as a facilitator in the Skills domain. When participants reported not needing to know the patient's gender because this information would not have changed their intervention, we mapped the barrier to Beliefs about consequences domain..."* Reference for the TDF grid with the definitions: Atkins, L., Francis, J., Islam, R., O'Connor, D., Patey, A., Ivers, N., Foy, R., Duncan, E.M., Colquhoun, H., Grimshaw, J.M., Lawton, R., Michie, S., 2017. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implementation Science 12.. doi:10.1186/s13012-017-0605-9	p. 20

11. Sex and gender are routinely recorded and clinical practice, but they do not seem to be considered when administering care unless the patient is undergoing menopause. It is unclear how the authors are advocating this intervention enters clinical consideration. How can this intervention speak to other countries/areas outside Quebec?	Thank you. We agree and we have given your questions a lot of thought. We have added a point to the discussion about how considerations of sex and gender could be brought into wider contexts of clinical care, other than patients in menopause alone, as follows: "Finally, in spite of the low priority given to sex and gender by our participants, qualitative analysis demonstrated that opportunities already exist for integrating these considerations into practice, such as the routine documenting of the patient's sex. CPD strategies could make more of these opportunities (60). *For example, CPD activities could advocate for sex- and gender-equitable care when treating men and women for diabetes and depression. Indeed, specific attention could be given to diabetic foot care when treating men, while specific attention could be given to blood-glucose regulation and to family and lifestyle issues when treating women (7, 61)."* We have also addressed the question of generalizability, as follows: *"This intervention could be adapted to medical fields other than T2D and depression, and to other countries and areas outside French-speaking provinces of Canada. While many of the barriers participants mentioned were culture- and language-specific to the Quebec or francophone context, many other languages (e.g. Spanish, German, Italian, and Portuguese) also generalize everything to the masculine gender, suggesting shared linguistic barriers. However, each culture has highly specific sex and gender norms affecting physicians' clinical assumptions (62). Our qualitative results highlight the fact that CPD on sex and gender considerations must be tailored to specific cultural contexts (17) and incorporate sex- and gender-based analysis tools (63)."*	p. 24 p. 24
Reviewer 2		
12. The goal of the study was to show feasibility, which the authors have demonstrated through recruitment and fidelity of the intervention. The conclusions of the study state that the intervention "had a favourable	Thank you for this comment. We acknowledge this limitation in our study, and mentioned it in the discussion as follows:	

impact on health care professionals' intention to include sex and gender considerations in caring for individuals with T2D and depression". However, have the authors considered how the non-randomized nature of the study which included participant preferences as well as the lack of pre-intervention measures of intent could influence the findings?	"Our study has a few limitations… *Second, the fact that participants could choose which course to attend (according to conference guidelines), hence the non-randomized nature of the study, may have biased our feasibility findings*." We also re-phrased this sentence in the conclusion: "*The significant between-group difference in the intention scores suggests the intervention had a favorable impact on health professionals' intention to include sex and gender considerations when caring for their patients with T2D and depression. However, caution is required as this effect may be attributed to other sources given the non-randomised nature for our study. Future randomised controlled trials are needed to control for potential selection biases and confirm our results, accounting for barriers and facilitators in sex- and gender-adapted diabetes and depression care."*	p. 24 p. 25
13. There are multiple subgroup analyses and comparisons, which introduces the potential for a false positive. The authors state the effectiveness of the intervention was greatest among older and rural-based physicians (page 23) but have the authors considered that the study was not designed or powered to show this definitively?	We fully agree with the reviewer. The differences observed between younger and older or between rural and urban professionals were based on bivariate analyses. We replaced the sentence in the Discussion with the following: *"Furthermore, bivariate analyses of the between-group difference in intention scores yielded significant results in older, but not younger, participants and in those practising in rural area."*	p. 23
14. The terms and definitions sex and gender appear to be used interchangeably at times in the manuscript. For example, page 8 "two males and one female" patient partners, page 10, line 46 "efforts were made to equally divide groups regrading number and sex of participants", line 53 "same two physicians (one male, assigned to the control group, and one female". Do the authors mean to identify the sex (rather than the gender identity)? Table 1 – should it be gender identity instead of sex? Similarly, page 23 line 18 "gender considerations" but then discuss female sexual health and cardiovascular risk, which are sex-based considerations.	Thank you, we have revised the manuscript to make sure our references to sex and gender are consistent throughout. We know from the literature* that both sex and gender influence disease manifestation, progression and treatment and we now refer to each of the terms based on the references used to explain the issues. For everything directly related to our study we made the change to consistently use the term "gender" which includes:	

	-gender identity [the related question on the study questionnaire was: Circle: I am: Man – Woman – Other] -gender norms/roles As such, we will refer to “gender” instead of “gender identity” specifically. We made the change to Table 1 to refer to “gender” with categories of women and men. *Available literature: Mauvais-Jarvis F, Bairey Merz N, Barnes PJ, Brinton RD, Carrero J-J, DeMeo DL, et al. Sex and gender: modifiers of health, disease, and medicine. The Lancet. 2020;396(10250):565-82. Oliffe JL, Greaves L. Designing and conducting gender, sex, and health research. Thousand Oaks: Sage Publications, Inc; 2012	p. 14
15. Have the authors considered that **the participants’ preference was taken into account in assignation to a group may have influenced the results due to bias?** Similarly, could the gender identity of the group leader have influenced the participants’ stated intentions?	Thank you for pointing this out. We agree, and we have added this to our limitations section. *“Second, the fact that participants could choose which course to attend (according to conference guidelines), and hence the non-randomized nature of the study, may have biased our feasibility findings. Also, the training was given by teachers of different gender for the intervention and control groups (a woman in the intervention group and a man in the control group). As a bias could have been introduced owing to differences in communication styles between men and women, the teaching teams practised giving the courses several times to ensure that teaching methods were equivalent. In addition, we ensured the teachers stayed with their respective groups for the six data collections.”*	p. 24
16. Are there studies that show stated intentions as measured by the CPD reaction questionnaire translate into clinical practice changes?	This is an excellent question and this is much debated in our field. We agree that intention is limited as a proxy for behavior, and that interventions based on the determinants of intention as dependent variables must also integrate methods to close the intention-behaviour gap. However, most CPD organizations find it more acceptable and feasible to assess behavioral intention than tracking the actual clinical behavior at stake. This	

	is not perfect but may, in this real world of CPD, be a mid-way agreement. In response we have added to our Introduction as follows: "…will help address the inequities between men and women. *Courses should be informed by theory-based factors known to influence the adoption of a given behaviour. Although one of several other factors influencing behaviour change, such as organizational constraints, intention is considered an acceptable proxy. Indeed, according to Godin's integrated model for health professional behaviour change, behavioral intention is the central influencing factor on behaviour adoption. In turn, this intention is under the influence of a number of other socio-cognitive factors (26).* We aimed to assess… *"* We also added the following to our Limitations section: "*While there is evidence that intention is an effective determinant for measuring behavior change (39), it is limited as a proxy. Finding other reliable measures of behaviour change is challenging (64).* However, identifying barriers and facilitators to change is a first step (64). Semi-structured group discussions using a clinical vignette have also been shown to contribute to clinical behaviour change (64). *Methods such as audit and feedback, as well as "commitment to change statements" could reduce the intention-behaviour gap and strengthen the understanding of clinical changes following CPD activities (65, 66)."*	p. 8 p. 25

VERSION 2 – REVIEW

REVIEWER	Alva, Maria Georgetown University, Massive Data Institute
REVIEW RETURNED	11-Jan-2022

GENERAL COMMENTS	While the authors now justify the importance of sex and gender in clinical practice, it is hard to imagine how a small increase in the reported likelihood of including sex/gender considerations would translate into action. The authors should provide/include concrete recommendations on what practitioners should do. That would be a different stand-alone paper targeted to a clinical audience.

	It is telling that while more than 3,000 practitioners were invited and only 127 accepted/consented. This appears to be an indication of the lack of interest from practitioners on the topic. While the authors answered many of my questions, the experimental design is still poorly discussed. The authors list in the paper now the many limitations: lack of baseline data, no RCT, etc. Having to gather consent forms or not being able to exclude people from courses are not strong justifications for these caveats. There are other designs that take into account these challenges like encouragement designs with double non-compliance. As a minor point, because the study was not an RCT, the authors should use innovation group, rather than treatment group to avoif confusion. I am still unclear on what is part of the intervention. The authors state in places that it is the same for all consenting 127 participants (intervention = 49 and control = 78). I am also unclear on how participants could have been blinded to the intervention and the control group. How could they avoid people from talking to each other? If both content and teacher differed between innovation and control group, how could we disentangle both effects? Ideally one would like to know (in aggregate form) how the 3,000+ non-participants and specifically the 51 who signed up but choose not to participate in the study were any different than those who did participate. Males, females? Organization type, etc.

REVIEWER	Ahmed, Sofia University of Calgary
REVIEW RETURNED	13-Dec-2021

GENERAL COMMENTS	The authors have done an excellent job responding to my queries. Congratulations on an important paper.

VERSION 2 – AUTHOR RESPONSE

REVIEWER 2

The authors have done an excellent job responding to my queries. Congratulations on an important paper.

Answer:

Thank you very much.

REVIEWER 1

1. While the authors now justify the importance of sex and gender in clinical practice, it is hard to imagine how a small increase in the reported likelihood of including sex/gender considerations would translate into action. The authors should provide/include concrete recommendations on what practitioners should do. That would be a different stand-alone paper targeted to a clinical audience.

It is telling that while more than 3,000 practitioners were invited and only 127 accepted/consented. This appears to be an indication of the lack of interest from practitioners on the topic.

Answer:

Thank you for the suggestion. The editor has suggested that we focus on describing the study in as clear a way as possible as well as noting limitations rather than making concrete recommendations.

As for the small number of practitioners who consented to participate in the study, this was not an RCT, but a controlled exploratory pragmatic study. We believe the response level may not be the main indicator of the importance of the topic, which had been suggested to us by the Médecins francophone du Canada (CPD provider) in response to physicians' stating that this was a gap in their education. Second, we started with the perhaps over-generous figure of n=3344. Most of these eligible practitioners were invited via email posters, or conference calendar invitations (we added this detail to our participant flow chart) as this is the usual way to invite potential participants to CPD activities. Our intent was to be integrated into the usual flow of invitations for large CPD gatherings such as those of our main CPD partner (Médecins Francophone du Canada). It is possible that health professionals, first, did not plan to attend any CPD activities at all. The target list included all members of the organisations we worked with. Also, it may be that those who received the invitation to our workshops may have simply not looked at them. Lastly, it is also important to know that in such a CPD context, attendance at interactive workshops is set to a maximum of participants to respect the room capacity. Also, it is well known that using passive strategies such as emails and advertising are not the most effective strategies for recruiting healthcare professionals in research (Parkinson A et al. Aust J Prim Health. 2015;21(2):254-8; Murphy C. C. et al. BMC Med Res Methodol, 2020 May 19;20(1):123).

In our study, for example, at one site (Montfort Hospital), the invitation was sent to nearly 2000 employees. These included 322 physicians as well as other healthcare professionals and staff (e.g. administrative staff). Thus, as with our other CPD invitation, many who got the invitation may not have been planning to attend any CPD activities at all. Also, some of them may not have been interested as they were not involved in treating patients with diabetes and depression. Also, in Médecins Francophones du Canada's conferences, all registered participants were invited to attend, as the CPD courses were announced in the conference calendar. However, these conferences covered a large range of medical topics, and many were taking place simultaneously.

In smaller sites, such as the Colloque francophone de médecine de Moncton, U-FMG Laurier and U-FMG Saint-François d'Assise, where potential eligible participants were approached more directly by the research team and where no other activities were happening at the same time, the participation rate was much higher, around 43-57%.

Very respectfully, we consider that our participation rate reflects the way CPD activities are publicized in large organizations and indeed highlights the pragmatic nature of the study.

2. While the authors answered many of my questions, the experimental design is still poorly discussed. The authors list in the paper now the many limitations: lack of baseline data, no RCT, etc. Having to gather consent forms or not being able to exclude people from courses are not strong justifications for these caveats. There are other designs that take into account these challenges like encouragement designs with double non-compliance.

Answer:

Our intention throughout this study was to integrate our research process as much as possible into the operations of a large CPD organization. We were geared towards a pragmatic approach that could be sustainable in the long term. In our context, where we needed a high level of flexibility, we discussed with our main CPD partner, Médecins Francophones du Canada, and chose a controlled trial with post intervention measure they felt it was more feasible to partner with us with this design. We agree that a different design might have overcome our limitations. However, this was an exploratory and pragmatic study, and we believe that despite limitations, we were able to assess the feasibility of the study and demonstrate that a CPD training on diabetes and depression that integrate sex and gender considerations may prompt health professionals to modify their care. We agree that future studies should take into consideration the limitations and challenges we faced in this study, and balance them against what real world CPD organizations are capable of.

3. As a minor point, because the study was not an RCT, the authors should use innovation group, rather than treatment group to avoid confusion.

Answer:

Thank you for this suggestion, we have replaced “intervention group” with “innovation group” throughout.

4. I am still unclear on what is part of the intervention. The authors state in places that it is the same for all consenting 127 participants (intervention = 49 and control = 78).

Answer:

The intervention was the same for all consenting participants (a diabetes and depression course), with the only exception of the teacher being different, and the addition of sex and gender content in the innovation group.

We added clarification in the Methods (Innovation) section:

“The original course, a 1-hour classroom-based activity, describes links between T2D and depression, reviews CANMAT 2016 Depression Guidelines and reviews pharmacological and non-pharmacological treatment of T2D and depression. This original course was used in the control group. Participants in the innovation group attended the same course but adapted to integrate sex- and gender-specific content including..”

5. I am also unclear on how participants could have been blinded to the intervention and the control group. How could they avoid people from talking to each other?

Answer:

In each site, both CPD courses (innovation group and control group) were given simultaneously in different rooms. Right after the training and while still in the room, participants were asked to complete the post intervention questionnaire. Therefore there was no communication between the groups.

We clarified this in the Methods (Innovation section) as follows:

"…the participants registered in one group or the other, both groups being blinded to the innovation and control group. Thus participants entered the classroom for whichever course they signed up for. There was no communication between these groups, as the two courses were given simultaneously. Participants had all received the same invitation to attend a course on T2D and depression. There was no mention of sex and gender content before participants entered the room."

6. If both content and teacher differed between innovation and control group, how could we disentangle both effects?

Answer:

Authors of studies that use active comparison conditions strive to match the two conditions carefully. The main difference between the two conditions in our case (apart from content, the effect we were interested in) was that the teachers were different. As we mentioned in the limitations of our revised version, the different teachers could indeed have affected our results. But blinding the groups to each other was only possible by giving the courses simultaneously, so we had no choice but to have 2 teachers. To minimize this difference, both teachers were involved in the development of the courses in collaboration with team members and patients partners.

Also, we would like to respectfully remind you that we were partnering with a large CPD organization with whom we discussed what was possible given their practical real-world constraints. We also mentioned in the limitations section of the manuscript that we had the teaching teams practise giving the courses several times to ensure that teaching methods were equivalent and used the same timeframe.

7. Ideally one would like to know (in aggregate form) how the 3,000+ non-participants and specifically the 51 who signed up but choose not to participate in the study were any different than those who did participate. Males, females? Organization type, etc.

Answer:

Unfortunately, we do not have information on the non-participants in this study, as they did not attend the courses. As previously mentioned, this was a pragmatic controlled trial with a high level of flexibility.

As illustrated in the flow chart (Figure 1), we excluded 32 participants (26 in control group and 6 in the innovation group), since they did not agree to sign the consent form.

However, since they completed the study questionnaire, we do have information on these 32 participants.

We performed sensitivity analyses including the 159 participants (127 who completed the questionnaire and signed consent form + 32 who completed questionnaire without signing the consent form) and found similar results. Including the 159 participants, 63.2% were under the age of 44, 80.0% were women, 70.2% were practising in French, 71.7% practised in the province of Quebec and 86.7% practised in an urban environment.

We also performed bivariate and multivariate analysis including the 159 participants and found results similar to those presented in the manuscript (data not shown).