PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Relapse prevention group therapy via video-conferencing for substance use disorder: protocol for a multicentre randomised controlled trial in Indonesia
AUTHORS	Yamada, Chika; Siste, Kristiana; Hanafi, Enjeline; Ophinni, Youdiil; Beatrice, Evania; Rafelia, Vania; Alison, Peter; Prabowo, Albert; Shinozaki, Tomohiro; Matsumoto, Toshihiko; Sakamoto, Ryota

VERSION 1 – REVIEW

REVIEWER	Greenblatt, Aaron University of Maryland School of Medicine, Psychiatry
REVIEW RETURNED	16-Mar-2021

GENERAL COMMENTS	Thanks for sharing this interesting protocol. I appreciate the hybrid model of implementation research and effectiveness, as well as the qualitative and quantitative outcomes. Questions: #) What are the demographic characteristics of the recruiting sites? Mostly urban, mostly rural? #) Please clarify: will patients in a given group all be from the same recruitment site? #) Primary outcome: Increase in percent days of abstinence from primarily used substance in the past 30 days--I am a little concerned that given that inclusion criteria simply require people to have used their problem substance once in the past 30 days that we may not be selecting for patients who are most likely to benefit from this intervention. #) Limitations that may be worth mentioning: 1) will participating in a group conducted in Bahasa Indonesia be a linguistic challenge for some patients? As a non-Indonesian reading this protocol I simply don't know enough about local linguistic features to understand if this is a barrier. 2) It is difficult to do a TLFB for a full 30 days; it would be ideal to do more frequent follow up for assessment than at T1, T2, T3, especially given that the intervention lasts a full 3 months. 3) I am concerned that the effect size of this psychosocial intervention will pale in comparison to the effect size of medication for opioid use disorder for patient participants receiving MMT. The population to be studied will be quite heterogeneous, and the Matrix model has its strongest evidence base for stimulant use disorder. 4) Does a WhatsApp group provide adequate privacy for this social setting? It means that each group participant will be disclosing personal contact information to other participants.

REVIEWER	Gallassi, Andrea Universidade de Brasilia, Programa de Pós-Graduação em Ciências e Tecnologias em Saúde (PPGCTS)
REVIEW RETURNED	18-May-2021

GENERAL COMMENTS	Lines 203, 204: The inclusion criteria 2 and 3 should be revised. If you have as an inclusion criteria people who have used primarily-used substance for at least one day in the last year, it doesn't seem appropriate to compare the results with someone who has used it for a long period in the last year. The authors should adjust the inclusion criteria, for example, at least one day in the last 30 days. Lines 213 to 215: Why was the urine test collected only after the participation of the program? Isn't there a urine sample collected before, in the baseline, to be able to compare and make the double check between the toxicological test and what the patients report? Lines 218 to 226: It is not clear whether the participants also were blinded. They should ensure a minimum of female participants(20% or 30% of the sample), to be able to analyse differences between gender.

REVIEWER	Koski-Jännes, A University of Tampere, Dept. of Social Sciences
REVIEW RETURNED	20-May-2021

GENERAL COMMENTS	This randomized controlled study aims to assess the effectiveness of a relapse prevention group therapy (Indo-DARPP) via videoconferencing for substance use disorders in Indonesia. The 12-week treatment is planned to be delivered in a small group format using a workbook under the guidance of one facilitator and one co-facilitator. The experimental group receives both the Indo-DARPP program and treatment as usual (TAU) whereas the control group receives only treatment as usual (TAU). The participants (N=160) will be recruited through three hospitals, two primary healthcare centers, and three rehabilitation centers. The primary outcome will be the number of days abstinent from the primary substance of use in the past 30 days. The secondary outcomes include addiction severity, quality of life, motivation to change, psychiatric symptoms, cognitive functions, and stress coping mechanisms. After the initial assessment and 12-week treatment the participants participate in the post-treatment assessment in week 13 and follow-up assessment at 24 weeks. The plan to use video conferencing as the platform for treatment delivery is clearly a good idea in a widespread country like Indonesia. The reference list displays familiarity with related treatment research. The study protocol is mostly well thought out and written and it adheres closely to the SPIRIT checklist. There are, however, some points in the plan that could be reconsidered: 1) The third inclusion criterion on p. 11 ”have used primarily-used substance for at least one day in the past year” looked rather strange. So loose a criterion does not seem to make sense in a substance use treatment study. 2) Only one full-day training of trainers looks too brief to be able to effectively cover all the items listed on p. 14: basic knowledge of SUD treatment, Indo-DARP contents, video demonstration, hands-on role-play, discussion of difficult cases, tele-Indo-DARPP, and study-related quality control. It is recommendable to extend the training to allow enough time for the participants to rehearse different aspects of the workbook and to discuss potential problems in treatment delivery. In fact, just learning the basics of e.g. relapse prevention or motivational interviewing would require much more

	training in case the treatment providers are not previously familiar with these treatment approaches. 3) To maintain treatment fidelity and quality control during the tele-Indo-DARPP sessions it is recommendable to observe more than just one random session at each site with a structured checklist. Since there will be 16 groups with twelve sessions each, it makes altogether 192 sessions. To secure treatment fidelity it would be recommendable to observe with a structured checklist randomly chosen 10–20%, i.e. 19–38 sessions. 4) One possibility could be to audiotape all the sessions (under the participants' concern). This would be particularly helpful in case the outcomes from each site would sharply differ from one another. It would also allow conducting further qualitative and quantitative research on the group interaction in this new type of treatment provision in Indonesia. 5) The session themes of the Indo-DARPP program or the contents of the workbook should be listed in the research protocol to allow the study to be repeated elsewhere. Another possibility would be to provide the net address of the workbook as a whole in the study protocol. 6)The authors rightly admit in the study limitations the potential problems created by the heterogenous control condition in different study sites. This will increase the variability of study results and possibly reduce the chances of finding significant differences between the control and experimental conditions. Furthermore, applying the new technique of videoconferencing particularly with participants from lower socioeconomic strata may create additional uncertainties and potential for early dropout. Therefore the assumed attrition proportion of 26% looks overly optimistic. Recruiting more than 160 participants in the study would be recommendable. 7) The reference list should include basic sources on relapse prevention (e.g. Marlatt & Gordon 1985) and motivational interviewing (e.g. Miller & Rollnick 2002; 2013) rather than just references of their short book reviews or overviews (references 30 and 31). 8) A third follow-up interview at six months would provide useful information on the stability of the study results.

VERSION 1 – AUTHOR RESPONSE

Responses to Reviewer 1

Thanks for sharing this interesting protocol. I appreciate the hybrid model of implementation research and effectiveness, as well as the qualitative and quantitative outcomes.

Questions:

#) What are the demographic characteristics of the recruiting sites? Mostly urban, mostly rural?

We appreciate the Reviewer for the important questions. All recruiting sites are located within the capital or big cities within each province, as seen in **Table 1**: six in Jakarta which is the national capital and the most densely populated city in the country, one in Tangerang which is a large city in the province of Banten, and one in Banda Aceh which is the capital of the Aceh province. Thus, the

recruiting sites themselves are classified as urban settings. However, the recruitment process will include online social media of each site, whose coverage is nationwide particularly for the rehabilitation centers, and we expect recruitment of participants from anywhere in Indonesia and not limited to the physical scope of each site. The detailed demographic characteristics of participants will be described in baseline assessment. We have added a clarification in the manuscript.

Methods: Participants and settings *(page 10, line 194 – page 11, line 197)*

Although all facilities are located in urban settings, the recruitment process will include online social media of each site, whose coverage is nationwide particularly for the rehabilitation centers, and we expect recruitment of participants from anywhere in Indonesia and not limited to the physical scope of each site.

#) Please clarify: will patients in a given group all be from the same recruitment site?

We thank the reviewer for pointing out the ambiguity. All patients in a given Indo-DARPP group will be from the same recruitment site.

Methods: Randomisation and blinding *(page 12, line 238-239)*

All participants in a given tele-Indo-DARPP group will be from the same recruitment site.

#) Primary outcome: Increase in percent days of abstinence from primarily used substance in the past 30 days--I am a little concerned that given that inclusion criteria simply require people to have used their problem substance once in the past 30 days that we may not be selecting for patients who are most likely to benefit from this intervention.

We fully agree with the Reviewer's comment and share the same concern. Previous studies (e.g. [1]) indicated that CBT or MI-based therapy would be more effective for participants who reported heavier substance use at baseline. Thus, stricting inclusion criteria to people who used substances in the past 30 days might increase the possibility to detect treatment effects. At the same time, however, we would like to maintain broader inclusion criteria, i.e. having used primary substance at least once in the past one year. We have two reasons for this. First, proportional hazard models showed that the probability of relapse remains high before achieving one year of abstinence, and only declined substantially after 16 months [2,3]. We thus believe it is important to include patients who haven't achieved one-year abstinence. Second, in this effectiveness study, we design eligibility criteria to fall into a more pragmatic side on the explanatory-pragmatic continuum [4], so that the study participants would more closely represent a population seen in real-world Indonesian clinical practice. Based on our experience at collaborating clinical sites, patients under treatment include people who are abstinent for more than 30 days but still experience cravings, and they indeed show interest in receiving relapse prevention psychotherapy.

Nevertheless, we realised that it is of interest to investigate the modification effect of substance use at baseline on treatment effectiveness, as Rosenblum et al [1] did. We thus explicitly wrote this analysis plan into our protocol.

Methods: Participants and setting *(page 11, line 216 - page 12, line 223)*

We set a broad inclusion criterion, i.e. substance use in the past one year, for two reasons. First, proportional hazard models showed that the probability of relapse remains high before achieving one year of abstinence, and only declined substantially after 16 months. Thus, it is clinically important to

examine treatment efficacy on people who have not achieved one-year abstinence. Secondly, in this pragmatic effectiveness study, we design eligibility criteria to more closely represent a population seen in real-world Indonesian clinical practice. Indeed, patients treated at the collaborating clinical sites include people who have been abstinent for more than one month, but still experience cravings and tendency to relapse.

Methods: Statistical analysis: Subgroup analysis *(page 23, line 469 - 472)*

Specifically, based on previous studies which showed that treatment effectiveness varied depending on baseline severity levels, we hypothesised that participants assigned to tele-Indo-DARPP with more severe level of substance use at T1 would report more increased abstinent days at T2, T3, and T4.

#) Limitations that may be worth mentioning:

1) will participating in a group conducted in Bahasa Indonesia be a linguistic challenge for some patients? As a non-Indonesian reading this protocol I simply don't know enough about local linguistic features to understand if this is a barrier.

We thank the reviewer for raising an important issue. The group psychotherapy will be only provided in the Indonesian language. A myriad of local languages can be found across the Indonesian archipelago, and while Bahasa Indonesia is the official national language, recent census showed that Bahasa speakers only account for more or less 80% of the nation's population. Older people residing in non-urban areas are also more likely to be not proficient in Bahasa, even though national data on substance users indicated a low proportion of older people—e.g., only 2.9% of people under drug rehabilitation service were aged >45 in a 2019 survey.

As suggested by the Reviewer, we acknowledged as a limitation the fact that this study result will not be generalisable to people with low language proficiency in **Discussion**.

Discussion *(page 27, line 570-572)*

Group psychotherapy provided in the Indonesian language would induce low external validity to people with limited proficiency in the language.

2) It is difficult to do a TLFB for a full 30 days; it would be ideal to do more frequent follow up for assessment than at T1, T2, T3, especially given that the intervention lasts a full 3 months.

We thank the Reviewer for this suggestion. We apologize for the insufficient explanation in the protocol, and would like to clarify that TLFB will be done on a weekly-basis within the 28 days period of assessment to reduce recall bias. We would like to employ TLFB for a one-month period for our primary outcome, because it has been widely used in studies with similar outcomes of self-reported substance use [5–8], and also because we found this method to be feasible in our pilot study.

Regarding the frequency of assessments, we would like to assess before and after the completion of the treatment, instead of in the middle of treatment, and thus believe that it is appropriate to assess at T1 and T2. Regarding timing and frequency of follow-up assessments, we believe that the assessment time point of three month after the treatment (T3) is close enough to T2 and we do not think we need to add further assessments between them. Nevertheless, as suggested

by another Reviewer, we decided to add another assessment one year after the treatment completion (i.e., T4) to see the durability of the treatment effect as a relapse prevention.

Methods: Primary outcome:Abstinence from primarily used substance *(page 16, line 323)*

*Use of primarily used substance each day (yes/no) for 28 days will be retrospectively interviewed on a weekly-basis using the timeline followback (TLFB) method (**Table 2**)...*

3) I am concerned that the effect size of this psychosocial intervention will pale in comparison to the effect size of medication for opioid use disorder for patient participants receiving MMT. The population to be studied will be quite heterogeneous, and the Matrix model has its strongest evidence base for stimulant use disorder.

Thank you for raising the concern about sample size. We agree that implementing this psychosocial intervention to a population, including those who are receiving MMT, might not achieve an effect size of 0.59. We thus proceeded to employ a more conservative estimate: an effect size of 0.5. Furthermore, we increased the expected attrition rate from 26% to 30%. Given these conditions, sample size was increased from 160 to 220. We proceeded to revise the sample size calculation in **Methods.**

Sample size *(page 21, line 429 - 431)*

Sample size was calculated for the primary outcome to detect a medium effect size of d = 0.50, which is slightly more modest than a previous study examining the efficacy of telemedicine for people with SUD in an LIMC (d = 0.59). Using α = 0.05 and power = 0.80, a simple t-test requires n = 64 per arm. We estimated the design effect of clustering within Indo-DARPP group, using the formula, D = 1 + (m −1) ρ, assuming intraclass correlation within Indo-DARPP groups or ρ = 0.05, and group size or m = 5, which yielded the design effect or D = 1.2. Multiplying the result of simple calculation by the design effect, the minimal number of participants in data analysis was 77 per arm. Assuming attrition proportion = 30% which is more conservative than a previous similar study (26%), the sample size for enrolment was set as 110 per arm, or 220 in total.

4) Does a WhatsApp group provide adequate privacy for this social setting? It means that each group participant will be disclosing personal contact information to other participants.

We thank the reviewer for raising an important point. During our pilot study, we found that creating a virtual group chat with CBT participants and their providers as members added positive influence into the group therapy, by establishing a supportive environment and friendly atmosphere. In post-study feedback sessions, pilot participants also felt encouraged by communicating in a group chat thanks to the sense of camaraderie and the overall convenience in joining the therapy. Some were even sharing words of encouragement on almost a daily basis. We are hoping that such experience can also be found in the actual study, and enhance the retention rate of CBT participants. Basic rules within the group chat (e.g., to keep social manners, and to not share information or group therapy URLs outside of the group) are properly explained and enforced throughout the study period. As for privacy concerns, WhatsApp employs end-to-end encryption which should ideally deter unwanted third-party intrusions.

Nevertheless, involvement in WhatsApp group chat is entirely voluntary; the risk of disclosing personal contacts by joining the group is properly informed beforehand. Participants who opt to not join the group chat would not be excluded from the Indo-DARPP therapy, and communication will be done via personal messages. We further elaborate in **Methods.**

Methods: Intervention via video conference: tele-Indo-DARPP *(page 14, line 274 - 276)*

URLs for video conferences will be informed weekly by the research team to five Indo-DARPP participants and two providers via online group chat for participants who agreed to share their contacts. Those who declined to share will be notified via personal messages.

Responses to Reviewer 2

Lines 203, 204: The inclusion criteria 2 and 3 should be revised. If you have as an inclusion criteria people who have used primarily-used substance for at least one day in the last year, it doesn't seem appropriate to compare the results with someone who has used it for a long period in the last year. The authors should adjust the inclusion criteria, for example, at least one day in the last 30 days.

We fully agree with the Reviewer's comment and share the same concern. Previous studies (e.g. [1]) indicated that CBT or MI-based therapy would be more effective for participants who reported heavier substance use at baseline. Thus, stricting inclusion criteria to people who have used substances for a long period in the past year, or have used substances in the past 30 days, will increase the homogeneity of the participants, thus might increase the possibility to detect treatment effects. At the same time, however, we would like to maintain broader inclusion criteria, i.e. having used primary substance at least once in the past one year. We have two reasons for this. First, proportional hazard models showed that the probability of relapse remains high before achieving one year of abstinence, and only declined substantially after 16 months [2,3]. We thus believe it is important to include patients who haven't achieved one-year abstinence. Second, in this effectiveness study, we design eligibility criteria to fall into a more pragmatic side on the explanatory-pragmatic continuum [4], so that the study participants would more closely represent a population seen in real-world Indonesian clinical practice. Based on our experience at collaborating clinical sites, patients are heterogeneous in terms of their current substance use status; those under treatment include people who are abstinent for more than 30 days but still experience cravings, and they indeed show interest in receiving relapse prevention psychotherapy.

Nevertheless, we realised that it is of interest to investigate the modification effect of substance use at baseline on treatment effectiveness, as Rosenblum et al [1] did. We thus explicitly wrote this analysis plan into our protocol. Incorporating advice from other Reviewers, we also increased the minimum sample size in this study plan from 160 to 220, which hopefully may increase the chance to detect treatment effects among heterogeneous participants.

Methods: Participants and setting *(page 11, line 216 – page 12, line 223)*

We set a broad inclusion criterion, i.e. substance use in the past one year, for two reasons. First, proportional hazard models showed that the probability of relapse remains high before achieving one year of abstinence, and only declined substantially after 16 months. Thus, it is clinically important to examine treatment efficacy on people who have not achieved one-year abstinence. Secondly, in this pragmatic effectiveness study, we design eligibility criteria to more closely represent a population seen in real-world Indonesian clinical practice. Indeed, patients treated at the collaborating clinical sites include people who have been abstinent for more than one month, but still experience cravings and tendency to relapse.

Methods: Statistical analysis: Subgroup analysis *(page 23, line 469 - 472)*

Specifically, based on previous studies which showed that treatment effectiveness varied depending on baseline severity levels, we hypothesised that participants assigned to tele-Indo-DARPP with more severe level of substance use at T1 would report more increased abstinent days at T2, T3, and T4.

Lines 213 to 215: Why was the urine test collected only after the participation of the program? Isn't there a urine sample collected before, in the baseline, to be able to compare and make the double check between the toxicological test and what the patients report?

We appreciate the Reviewer for asking an important question. In this study, we regard the purpose of the urine test merely as supplemental validation of self-reported drug use during the period while the drug metabolite is detectable in urine, instead of surrogate for primary outcome nor rigorous, full validation of self-reported drug use. To act as the latter, we agree with the Reviewer that it is better to collect urine samples both before and after the intervention. Moreover, for better validation, ideally we would collect urine samples frequently (e.g. every 3 days) for one-month duration at each assessment point, since detection period for some drug metabolites in urine are only up to 3 days after the occurrence of drug use, or else to analyze other longer term specimens such as hair.

The reason for us to not conduct a full validation is due to feasibility issues in part of the participants. It is imperative to minimise burden for participants, especially those who reside in remote areas, to contact researchers in-person to submit urine samples. Our study aims to evaluate online therapy which can be provided no matter how inaccessible treatment facilities are from their residence. In line with this, all data collection except for urine collection will be done online. Increased burden might cause drop-out from the study, especially among participants who deemed most benefited from online therapy. Thus, to find a middle-ground, and in line with the nature of our study as a pragmatic study, we plan to conduct a urine test only once at our primary endpoint, T2, and use the objective data as a partial validation to corroborate the self-reported drug use data. This strategy, i.e. objective biospecimen analysis only at posttreatment time point for confirmation purposes, has also been done in prior similar studies or described in protocols [9–11]. We added this explanation in **Methods** and **Discussion**.

Methods: Primary outcome: Abstinence from primarily used substance *(page 17, line 335 - 340)*

Urine tests in this study will only serve to corroborate the data of self-reported substance use at the primary endpoint (T2), and not as an objective surrogate of all self-reported substance use data at every time point. This was planned to improve feasibility for participants and minimise drop-out due to the burden of data collection (urine test needs in-person assessment, unlike all other measurements in this study), especially among participants who reside in remote areas deemed most benefited from online therapy.

Discussion *(page 27, line 560 - 563)*

Urine tests will be done to corroborate subjective data, but this is not a full validation as it is only done once to represent substance-detectable period within a 28-days period, and only at T2. This in turn was planned to improve feasibility for participants and reduce drop-out risk, as all other data collection will be done online except only for urine tests.

Lines 218 to 226: It is not clear whether the participants also were blinded.

Thank you for pointing out the ambiguity. As the nature of the intervention is a behavioral therapy, it would be impossible to blind either the participants or the therapists to the group allocation. The resulting risk of bias and the impossibility to control for placebo effects in CBT trials are already addressed elsewhere [12]. Nevertheless, we have applied blinding to both randomiser and data

collectors in all timepoints. We proceeded to clearly state that the participants and providers are not blinded in **Methods**.

Methods: Randomisation and blinding *(page 13, line 243 - 244)*

Participants and treatment providers are not blinded as the intervention is a psychotherapy.

They should ensure a minimum of female participants(20% or 30% of the sample), to be able to analyse differences between gender.

We appreciate the Reviewer's advice. Similar to most data elsewhere, the proportion of female participants in SUD treatment in Indonesia is much lower than that of male—a 2019 national report shows the prevalence comparison between female to male who have ever used substances as 2.3% vs 6.5% [13]. We thus plan to conduct subgroup analysis only when we are able to recruit a sufficient number of female patients which enables us to conduct meaningful comparison. This is also the case for other subgroup categories, such as types or primarily used substance, concurrent use of other SUD treatment, or baseline URICA readiness score. We proceeded to mention this in **Methods**. Also, in line with advice from other Reviewers, we have decided to increase the sample size in this study from 160 to 220, which hopefully will allow us to do such subgroup analyses.

Methods, Statistical analysis, Subgroup Analysis (page 23, line 473-474)

These subgroup analyses will be conducted only when sufficient numbers of participants are available for certain categories to draw meaningful comparisons.

Responses to Reviewer 3

This randomized controlled study aims to assess the effectiveness of a relapse prevention group therapy (Indo-DARPP) via videoconferencing for substance use disorders in Indonesia. The 12-week treatment is planned to be delivered in a small group format using a workbook under the guidance of one facilitator and one co-facilitator. The experimental group receives both the Indo-DARPP program and treatment as usual (TAU) whereas the control group receives only treatment as usual (TAU).

The participants (N=160) will be recruited through three hospitals, two primary healthcare centers, and three rehabilitation centers. The primary outcome will be the number of days abstinent from the primary substance of use in the past 30 days. The secondary outcomes include addiction severity, quality of life, motivation to change, psychiatric symptoms, cognitive functions, and stress coping mechanisms. After the initial assessment and 12-week treatment the participants participate in the post-treatment assessment in week 13 and follow-up assessment at 24 weeks.

The plan to use video conferencing as the platform for treatment delivery is clearly a good idea in a widespread country like Indonesia. The reference list displays familiarity with related treatment research. The study protocol is mostly well thought out and written and it adheres closely to the SPIRIT checklist.

We are grateful for the detailed observation and positive feedback. We appreciate the Reviewer's agreement to the idea of using video conferencing platforms as treatment delivery in the vast archipelago of Indonesia. Unfortunately, evidence for this is non-existent in the country, or even for any national standardised formal treatment targeted for people with SUD, which prompted us to plan this RCT.

There are, however, some points in the plan that could be reconsidered:

1) The third inclusion criterion on p. 11 "have used primarily-used substance for at least one day in the past year" looked rather strange. So loose a criterion does not seem to make sense in a substance use treatment study.

We share the same concern with the Reviewer and have spent a great amount of time scrutinising the inclusion criteria. Previous studies (e.g. [1]) indicated that CBT or MI-based therapy would be more effective for participants who reported heavier substance use at baseline. Thus, strictening inclusion criteria to people who have used substances for a long period in the past year, or have used substances in the past 30 days, will increase the homogeneity of the participants, thus might increase the possibility to detect treatment effects. At the same time, however, we would like to maintain broader inclusion criteria, i.e. having used primary substance at least once in the past one year. We have two reasons for this. First, proportional hazard models showed that the probability of relapse remains high before achieving one year of abstinence, and only declined substantially after 16 months [2,3]. Thus, as we will confirm the DSM-5 diagnosis in the first inclusion criterion, we then decided to include those at higher risk of relapse which means those that haven't achieved a full year of abstinence, and set a very loose threshold for this, i.e. those that still uses substance for at least a day in the past year. Such once-in-a-year criterion has actually been used in prior SUD treatment studies [14], and some did not even put current substance use as their inclusion criteria [1,15].

Secondly, in this effectiveness study, we design eligibility criteria to fall into a more pragmatic side on the explanatory-pragmatic continuum [4], so that the study participants would more closely represent a population seen in real-world Indonesian clinical practice. Based on our experience at collaborating clinical sites, patients are heterogeneous in terms of their current substance use status; those under treatment include people who are abstinent for more than 30 days but still experience cravings, and they indeed show interest in receiving relapse prevention psychotherapy. Nevertheless, we realised that it is of interest to investigate the modification effect of substance use at baseline on treatment effectiveness, as Rosenblum et al [1] did. We thus explicitly wrote this analysis plan into our protocol. Incorporating the other advice from the Reviewer, we also increased the minimum sample size in this study plan from 160 to 220, which hopefully may increase the chance to detect treatment effects among heterogeneous participants.

Methods: Participants and setting *(page 11, line 216 – page 12, line 217)*

We set a broad inclusion criterion, i.e. substance use in the past one year, for two reasons. First, proportional hazard models showed that the probability of relapse remains high before achieving one year of abstinence, and only declined substantially after 16 months. Thus, it is clinically important to examine treatment efficacy on people who have not achieved one-year abstinence. Secondly, in this pragmatic effectiveness study, we design eligibility criteria to more closely represent a population seen in real-world Indonesian clinical practice. Indeed, patients treated at the collaborating clinical

sites include people who have been abstinent for more than one month, but still experience cravings and tendency to relapse.

Methods: Statistical analysis: Subgroup analysis *(page 23, line 469 - 474)*

Specifically, based on previous studies which showed that treatment effectiveness varied depending on baseline severity levels, we hypothesised that participants assigned to tele-Indo-DARPP with more severe level of substance use at T1 would report more increased abstinent days at T2, T3, and T4.

2) Only one full-day training of trainers looks too brief to be able to effectively cover all the items listed on p. 14: basic knowledge of SUD treatment, Indo-DARP contents, video demonstration, hands-on role-play, discussion of difficult cases, tele-Indo-DARPP, and study-related quality control. It is recommendable to extend the training to allow enough time for the participants to rehearse different aspects of the workbook and to discuss potential problems in treatment delivery. In fact, just learning the basics of e.g. relapse prevention or motivational interviewing would require much more training in case the treatment providers are not previously familiar with these treatment approaches.

We appreciate the Reviewer for pointing out an important issue. After reconsideration of the training schedule, we have decided to double the amount of time to two full-day training. This change is also based on the government-accredited training for SMARPP in Japan—the original module where Indo-DARPP was adopted from—which was provided in two full-day training sessions. We have amended the manuscript in Methods.

Methods: Training and supervision *(page 15, line 299)*

Prior to recruitment, all providers will receive two full-day training online sessions...

3) To maintain treatment fidelity and quality control during the tele-Indo-DARPP sessions it is recommendable to observe more than just one random session at each site with a structured checklist. Since there will be 16 groups with twelve sessions each, it makes altogether 192 sessions. To secure treatment fidelity it would be recommendable to observe with a structured checklist randomly chosen 10–20%, i.e. 19–38 sessions.

We agree that we should monitor the treatment provision more closely, and we are thankful to the Reviewer for the calculation and practical recommendation. We thus proceeded to change the monitoring method to observe two sessions per each Indo-DARPP group (per wave per site), which constitute 16.7% of the all sessions. We have revised the manuscript in **Methods.**

Methods: Training and supervision *(page 15, line 304 – page 16, line 306)*

...addiction psychiatrists from the research team (KS and EH) will randomly select and observe at least two sessions per Indo-DARPP group, meaning at each wave at each site thus constituting 16.7% of the all sessions...

4) One possibility could be to audiotape all the sessions (under the participants' concern). This would be particularly helpful in case the outcomes from each site would sharply differ from one another. It

would also allow conducting further qualitative and quantitative research on the group interaction in this new type of treatment provision in Indonesia.

Thank you for the practical suggestion; we indeed were heavily considering recording every Indo-DARPP session for further qualitative review. However, we ended up deciding against it after conducting the pilot test, which we recorded under participants' consent. To our observation, when the recording was on, some participants were conscious about the fact that what they were saying was recorded and mentioned that they accidentally said something that they should have not. While these comments were said in jest (which indeed incited big laughs from all others) and in a friendly manner, we learnt that recording quite obviously affects the participants' thinking and behaviour. To facilitate honest disclosure of their thoughts, or at least provide a feeling of freedom for the participants, and keeping with our promise to them to ensure their privacy, we decided not to record sessions in RCT. Unfortunately, matters of illicit substance is heavily criminalised in Indonesia—traffickers punished with death penalty and lifetime incarceration, and whenever there is such incident it is recurrently aired in the news—and using illicit substance is still considered as a taboo subject not fitting to be talked about in public spaces. This may instill a feeling of fear for participants if we record the conversation. Furthermore, our site collaborators also include Puskesmas, which are government-based institutions, so by not recording the sessions we hopefully avoid enacting a punitive image in the participants' mind, or heavy surveillance usually associated with the government.

We appreciate the Reviewer's comment, and proceeded to mention our decision in Methods.

Methods: Data collection procedure (*page 21, line 422****)***

To facilitate honest disclosure from participants, we will not record any Indo-DARPP video conferencing sessions throughout the study.

5) The session themes of the Indo-DARPP program or the contents of the workbook should be listed in the research protocol to allow the study to be repeated elsewhere. Another possibility would be to provide the net address of the workbook as a whole in the study protocol.

Thank you and we apologise for not properly showing any contents of the workbook. We proceeded to include the Table of Contents of the Indo-DARPP workbook, translated into English, in **Supplementary file 3**. However, we avoided uploading the whole book to the internet, as the book will be used as an intervention in an ongoing trial and doing so may inadvertently contaminate the control participants. Nevertheless, we plan to distribute the workbook after the trial has ended and after we have properly published and disseminated the results.

As a side note, we are currently writing another paper describing the development of the Indo-DARPP module as well as its themes and contents, together with pilot study results and feedback. Hopefully if this development paper can be published, it may serve as a reference for other stakeholders to adapt the module contents into their own local contexts.

Methods: Development of Indo-DARPP *(page 14, line 263-265)*

Indo-DARPP is designed to be delivered in a small group format using a workbook (see ***Supplementary file 3*** *for table of contents of the workbook)...*

6)The authors rightly admit in the study limitations the potential problems created by the heterogenous control condition in different study sites. This will increase the variability of study results and possibly reduce the chances of finding significant differences between the control and experimental conditions. Furthermore, applying the new technique of videoconferencing particularly with participants from lower socioeconomic strata may create additional uncertainties and potential for early dropout. Therefore the assumed attrition proportion of 26% looks overly optimistic. Recruiting more than 160 participants in the study would be recommendable.

We are grateful to the Reviewer for raising the concern about sample size. We fully agree that the heterogeneity of TAU and possible dropout due to new techniques implemented as intervention should be considered. We thus proceeded to employ a more conservative expectation, which is an effect size of d=0.5, than the original plan (d=0.59). Furthermore, we increased the expected attrition rate from 26% to 30%. Given these conditions, sample size was increased from 160 to 220. We proceeded to revise the calculation in **Methods**.

Methods: Sample size *(page 21, line 429 - 438)*

Sample size was calculated for the primary outcome to detect a medium effect size of d = 0.50, which is slightly more modest than a previous study examining the efficacy of telemedicine for people with SUD in an LIMC (d = 0.59). Using α = 0.05 and power = 0.80, a simple t-test requires n = 64 per arm. We estimated the design effect of clustering within Indo-DARPP group, using the formula, D = 1 + (m −1) ρ, assuming intraclass correlation within Indo-DARPP groups or ρ = 0.05, and group size or m = 5, which yielded the design effect or D = 1.2. Multiplying the result of simple calculation by the design effect, the minimal number of participants in data analysis was 77 per arm. Assuming attrition proportion = 30% which is more conservative than a previous similar study (26%), the sample size for enrolment was set as 110 per arm, or 220 in total.

7) The reference list should include basic sources on relapse prevention (e.g. Marlatt & Gordon 1985) and motivational interviewing (e.g. Miller & Rollnick 2002; 2013) rather than just references of their short book reviews or overviews (references 30 and 31).

We thank the reviewer for kindly indicating fundamental sources of relapse prevention and motivational interviewing, and we apologise for missing to mention them. In accordance with the Reviewer advice, we have cited the following articles:

8) A third follow-up interview at six months would provide useful information on the stability of the study results.

We absolutely agree with the Reviewer that another assessment at a later time point would provide useful information on the stability of the results. As we stated earlier, prior knowledge indicates more or less one year of abstinence as a critical duration to maintain long-term abstinence. Thus, we decided to add another assessment of T4 at one year after the treatment ends. We have revised the **Methods** as well as **Figure 1 and 2** to reflect this addition. Although it would be ideal to also conduct assessment at six months post-treatment, we sincerely apologize that our humble research budget would not allow us to conduct two additional assessments.

Methods: Data collection procedure *(page 20, line 415 -416)*

Researchers blinded to the treatment allocation will collect data at three different time points: at baseline (week 0, T1), the week after the completion of treatment (week 13, T2), three months after the completion of treatment (week 24, T3), and twelve months after the completion of treatment (week 60, T4)

VERSION 2 – REVIEW

REVIEWER	Gallassi, Andrea Universidade de Brasilia, Programa de Pós-Graduação em Ciências e Tecnologias em Saúde (PPGCTS)
REVIEW RETURNED	29-Jun-2021

GENERAL COMMENTS	The authors have addressed the questions and suggestions raised by the reviewers. The protocol seems that will be very useful to cover who need treatment for substance use disorder but can't access; for that the research team needs to make sure that the language will be appropriate for all participants of the program.

REVIEWER	Koski-Jännes, A University of Tampere, Dept. of Social Sciences
REVIEW RETURNED	21-Jul-2021

GENERAL COMMENTS	The revisions to the former version are satisfactory. A description of how motivational interviewing is used in this mainly relapse prevention type of program would have still improved the plan.

VERSION 2 – AUTHOR RESPONSE

Responses to Reviewer 2

The authors have addressed the questions and suggestions raised by the reviewers. The protocol seems that will be very useful to cover who need treatment for substance use disorder but can't access; for that the research team needs to make sure that the language will be appropriate for all participants of the program.

We highly appreciate the Reviewer for the important feedback throughout the revision process, and for acknowledging the value of this program to reach the hard-to-reach population of people with SUD.

Regarding language; we share the Reviewer's concern. The workbook is written in Bahasa Indonesia, and in this trial we included proficiency in Bahasa Indonesia as an inclusion criterion. Nevertheless, we are hoping that the program can be implemented all over the Indonesian archipelago and language barriers will be one problem to overcome. While Bahasa Indonesia is the official national language, a recent census showed that Bahasa speakers only account for ~80% of the population. Older people residing in non-urban areas are also more likely to be not proficient in Bahasa.

We clarify the point as a study limitation in **Discussion**:

Discussion *(page 27, line 564)*

Psychotherapy will be provided in Bahasa Indonesia; hence its effectiveness would not be generalisable to people with limited proficiency in the language.

Also, we will include questions regarding communication and language barrier in the post-treatment qualitative interview with the participants.

Methods: Implementation outcomes: Feedback interviews *(page 20, line 398)*

Semi-structured interviews will be conducted with both participants and providers to assess the following: satisfaction with content quality, comprehensibility, technical experience regarding video-conferencing, comfort, module practicability, language barriers, and participants' perception of the credibility of providers. Interviews will be audio-recorded with the interviewees' consent.

Moreover, in the future, we are hoping to further develop Indo-DARPP into a more localised version, such as translating into major local languages in Indonesia (e.g., Javanese), and including more nuanced cultural references. This remains as an arbitrary plan, however, and thus is not mentioned in the current protocol manuscript.

Responses to Reviewer 3

The revisions to the former version are satisfactory. A description of how motivational interviewing is used in this mainly relapse prevention type of program would have still improved the plan.

As suggested by the Reviewer, we included a more elaborate description on how motivational interviewing (MI) is incorporated into the relapse prevention program.

Methods: Development of Indo-DARPP *(page 13, line 251)*

Elements of MI are incorporated in the earlier parts of the workbook in a form of open questions to assess participants' ambivalence and motivation to change.

Methods: Training and supervision *(page 15, line 298)*

Prior to recruitment, all providers will receive two full-day online training sessions on basic knowledge of SUD treatment, Indo-DARPP content, principles of MI (e.g., empathy, reflective listening, empowering affirmations), video demonstrations, hands-on role play, discussion of difficult cases, and study-related quality control

We highly appreciate the Reviewer for the kind expertise and detailed advice throughout the revision process.