PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Task-sharing with lay counsellors to deliver a stepped care intervention to improve depression, antiretroviral therapy adherence, and viral suppression in people living with HIV: study protocol for the TENDAI randomised controlled trial.
AUTHORS	Abas, Melanie; Mangezi, Walter; Nyamayaro, Primrose; Jopling, Rebecca; Bere, Tarisai; McKetchnie, Samantha M.; Goldsmith, Kimberley; Fitch, Calvin; Saruchera, Emily; Muronzie, Thabani; Gudyanga, Denford; Barrett, Barbara; Chibanda, D; Hakim, James; Safren, Steven; O'Cleirigh, Conall

VERSION 1 – REVIEW

REVIEWER	Tran, Cuc Division of HIV & Global Tuberculosis, U.S. Centers for Disease Control and Prevention
REVIEW RETURNED	29-Nov-2021

GENERAL COMMENTS	Nice study design. I look forward to seeing the outcomes from this study. The authors should provide additional details regarding compensation/incentives for the participants aside from transportation. If compensation is not provided, it should be clearly stated. The authors should address potential selection bias (compensation or no compensation may impact who is willing to participate), skewing the representativeness of the findings (e.g., vulnerable populations may have different needs.). The authors may want to address confounders associated to depression, and how it may impact the findings (meaning is did the intervention itself improve adherence or where there other factors that may have impacted adherence in positively or negatively?)

REVIEWER	Velloza, Jennifer University of Washington
REVIEW RETURNED	13-Dec-2021

GENERAL COMMENTS	This manuscripts describes a novel task-sharing intervention to improve both depression and viral suppression among people living with HIV. It is well-written and clear, and the study design, sample, and methods are appropriate for answering the questions of interest. I have minor edits and suggestions where additional clarity is needed. It should also be noted that I was assisted in this review by Tessa Concepcion, a PhD student at the University of Washington.

	1. Abstract: Instead of saying “meet criteria for depression”, please say something like probable depression, likely depression, elevated depressive symptoms, or depressive symptoms 2. Abstract: Define “persistent depression” 3. Strengths: The authors list a strength that this is an RCT to test an intervention to improve adherence and depression, but they are not exploring depression as a primary outcome. Please adjust this language accordingly. 4. Strengths: What does the phrase “cost-effectiveness under investigation” mean? 5. Strengths: “Stepped care intervention [can?] be delivered” 6. Introduction: In the opening paragraph, it would be helpful to add some literature on U=U and the importance of undetectable viral loads as not only an HIV treatment strategy but also an effective HIV prevention approach 7. Introduction: Please add a citation for the phrase “through interfering with the uptake of existing adherence support programs” 8. Introduction: Can the authors state that TENDAI is a culturally adapted CBT-AD intervention? 9. Methods: Please revise the phrase “if prescribed anti-depressants, be on a stable regimen”. 10. Methods: How are untreated or undertreated mental illness, advanced physical disease, and severe cognitive impairment being assessed? 11. Methods: What is meant by “HIV characteristics”? What “measures of socio-economic position” are being used? 12. Methods: Dried blood spot (DBS) [testing]. Please also provide additional detail about where DBS testing is done and thresholds for ART adherence assessment. 13. Methods: Are the RAs collecting data like alcohol and substance use information different from the Independent Assessor? 14. Methods: Add an apostrophe to “participants life goals” 15. Methods: How is “persistent depression” being assessed? 16. Methods: How far apart are intervention sessions? Are they done weekly? Monthly? 17. Methods: How will “diagnosis of depression” be made for participants in the control arm? Based on the MINI? The PHQ-9? 18. Methods: Please provide rationale for the use of the PHQ-9 as the instrument for assessing depressive symptoms in this cohort. 19. Methods: Please provide a bit more detail on the training of interventionists. Who is conducting the training? Will refresher training be done throughout study follow-up? 20. Methods: How is self-reported adherence being assessed? 21. Methods: Spell out “SAEs” the first time it is used 22. Methods – Confidentiality section: Capitalize the “P” in participants (“participants are referred to only by their identification number…”). 23. Methods – Blinding section: “The IA…and the lead study statistician [are] blinded” 24. Methods: What do the authors mean by following an intention to treat principle “as much as possible”? 25. Methods: Do the authors mean that they will explore moderation by “sex” or by gender? 26. Methods: The authors say “where these data are missing…they will be coded as not suppressed, otherwise the outcome will be left missing”. Can they clarify if all missing data is coded as not suppressed or missing?

	27. Methods: Please review the line spacing in the “Dissemination” section. 28. Trial status: Please also state where the trial currently is in terms of enrollment and data collection.

REVIEWER	McBain, Ryan RAND Corp
REVIEW RETURNED	31-Dec-2021

GENERAL COMMENTS	The authors provide a study protocol for the TENDAI randomized controlled trial, which provides a stepped care intervention for improving depression ART adherence and viral suppression among adults living with HIV in Zimbabwe. The protocol is generally well-written, following appropriate conventions for structure and formatting as well as completion of the SPIRIT checklist. I raise only a few points for the authors to consider, outlined below. 1. Based on the authors’ power calculation, they note 85% power to detect an absolute difference of 20% or more in achieving viral suppression. While I understand that pilot data have shown a shift of 25%, this nevertheless strikes me as a very large difference. Are there any other interventions beyond the pilot data that have identified shifts of this magnitude? If so, mentioning these and accompanying citations would be helpful. 2. For secondary and tertiary outcomes, do the authors plan to perform a correction for multiple tests? Either way, I would state this in the Statistical Methods section. 3. Spacing between lines in the “Dissemination” subsection appears to be different (wider) from the rest of the manuscript. 4. The authors mention a preference for reporting mean costs despite (likely) skewed data and cite Thompson and colleagues’ 2000 BMJ article. Could you elaborate more on this justification? (This appears to be an appeal to authority without actually providing the rationale.) 5. Presumably, the cost-effectiveness component will involve ICERs (incremental cost-effectiveness ratios); however, there is no mention of this. If this is correct, I would suggest adding text to this effect.

VERSION 1 – AUTHOR RESPONSE

Response to Reviewers
Reviewer 1: Dr. Cuc Tran

1. The authors should provide additional details regarding compensation/incentives for the participants aside from transportation. If compensation is not provided, it should be clearly stated. The authors should address potential selection bias (compensation or no compensation may impact who is willing to participate), skewing the representativeness of the findings (e.g., vulnerable populations may have different needs.).

As participants are going through the assessments, refreshments (water and biscuits) are provided. No additional compensation besides the transportation costs are provided. Although those in the EUC arm have fewer scheduled clinical sessions, the same total amount will be provided to participants attending all clinical sessions and research assessments in both arms. Reimbursement is given in line

with local standards for providing a brief amount of data at clinical sessions in Zimbabwe. It is not expected that this would skew the data.

	Control Arm	Intervention Arm
Baseline Session	1 session @ $6	1 session @ $6
Research Assessment	3 sessions @ $8	3 sessions @ $4
Clinical Session	2 sessions @ $8	6 sessions @ $4
Booster Clinical Session		1 session @ $4
Total	6 sessions @ $46	9 sessions @ $46

2. The authors may want to address confounders associated to depression, and how it may impact the findings (meaning is did the intervention itself improve adherence or were there other factors that may have impacted adherence in positively or negatively?)

We thank the reviewer for the thoughtful read of our manuscript. Our intervention is designed to reduce symptoms of depression and increase HIV medication adherence directly through the proximal effects of the specific treatment components. We hypothesize that these treatment related changes will impact HIV viral load more distally. This intervention is manipulated by its presence in the experimental condition and its absence in the control condition. We have planned mediation analyses to examine the relative contributions of treatment related changes in depression and in adherence on treatment related changes in HIV viral load. This will provide an empirical test of our conceptual model. We have referred to these mediation analyses in the revised manuscript but without explication due to space limitations.

Reviewer 2: Dr. Jennifer Velloza, University of Washington

1. Abstract: Instead of saying “meet criteria for depression”, please say something like probable depression, likely depression, elevated depressive symptoms, or depressive symptoms.

This has been changed to “probable depression” (Patient Health Questionnaire >10).

2. Abstract: Define “persistent depression”

This has been replaced with “Participants whose score on the PHQ-9 remains >10, and/or falls by less than 5-points, step up to a nurse-evaluation for possible antidepressant medication.”

3. Strengths: The authors list a strength that this is an RCT to test an intervention to improve adherence and depression, but they are not exploring depression as a primary outcome. Please adjust this language accordingly.

Depression is assessed as a key secondary outcome, with the trial powered to detect a difference in viral load as the primary outcome. The authors therefore are testing an intervention to improve both viral suppression and depression.

4. Strengths: What does the phrase “cost-effectiveness under investigation” mean?

This sentence has been revised to reflect the assessment of the cost-effectiveness of the TENDAI stepped care task-shifted intervention as a strength of this study.

5. Strengths: “Stepped care intervention [can?] be delivered”

This sentence has been revised to read “Stepped care intervention is delivered through task-shifting to non-specialist staff, allowing for future scale up.”

6. Introduction: In the opening paragraph, it would be helpful to add some literature on U=U and the importance of undetectable viral loads as not only an HIV treatment strategy but also an effective HIV prevention approach

A sentence in the importance of undetectable viral loads has now been included. We thank the reviewer for this comment.

7. Introduction: Please add a citation for the phrase “through interfering with the uptake of existing adherence support programs”

This sentence has now been revised.

8. Introduction: Can the authors state that TENDAI is a culturally adapted CBT-AD intervention?

TENDAI is not CBT-AD. Please see page 5. We removed the word CBT-AD and replaced it with “Recent reports, including from our team in Zimbabwe, support the acceptability and feasibility of culturally-adapted cognitive behavioural interventions for low resource settings.” Further at the end of the Introduction we state “TENDAI is derived from principles of problem-solving and psychoeducation for depression and adherence, and motivational interviewing.”

9. Methods: Please revise the phrase “if prescribed anti-depressants, be on a stable regimen”.

This sentence has been revised to “if prescribed anti-depressants, have been on a stable anti-depressant regimen for at least 2 months.”

10. Methods: How are untreated or undertreated mental illness, advanced physical disease, and severe cognitive impairment being assessed?

The exclusion criteria have been clarified in the manuscript:
Exclusion criteria: 1) unwilling or unable to provide informed consent; 2) major untreated or undertreated mental illness (e.g., untreated psychosis or mania, actively suicidal assessed using the MINI and P4 suicide screener), major or advanced physical disease (assessed using clinic records) or severe cognitive impairment (assessed using international HIV dementia scale) which would interfere with engagement in Stepped Care-AD.

11. Methods: What is meant by “HIV characteristics”? What “measures of socio-economic position” are being used?

The term HIV characteristic have been removed as this referred to key HIV markers including viral load and CD4 count. The measures of socio-economic position have been clarified: “measures of socio-economic position (employment status, educational history, and ownership of household assets).”

12. Methods: Dried blood spot (DBS) [testing]. Please also provide additional detail about where DBS testing is done and thresholds for ART adherence assessment.

DBS samples have been collected and are securely stored at the African Institute of Biomedical Science and Technology, Harare, Zimbabwe. We are currently liaising with partners at the University

of Cape Town to finalize the location for analysis. As the location is not yet finalized, we have not added details of the location and threshold used.

13. Methods: Are the RAs collecting data like alcohol and substance use information different from the Independent Assessor?

The Independent Assessor (IA) administers the PHQ-9, EQ-5D-3L, and self-report medication adherence measures. They do not ask questions about alcohol and substance use. RAs will collect self-report data including use of alcohol and substances.

14. Methods: Add an apostrophe to “participants life goals”

Thank you for this comment, this has now been revised.

15. Methods: How is “persistent depression” being assessed?

Persistent depression is assessed as depression score continuing above cut-off (PHQ-9 >10) or if they have less than a 5-point improvement in PHQ-9 score from Session 4. This has now been clarified in the manuscript.

16. Methods: How far apart are intervention sessions? Are they done weekly? Monthly?

Intervention sessions are delivered weekly. This has been clarified in the manuscript.

17. Methods: How will “diagnosis of depression” be made for participants in the control arm? Based on the MINI? The PHQ-9?

This sentence has been revised in the manuscript for clarity. “We will provide a letter for each participant communicating the patients PHQ-9 score to their HIV-care provider.”

18. Methods: Please provide rationale for the use of the PHQ-9 as the instrument for assessing depressive symptoms in this cohort.

The PHQ-9 has been validated for adults in a primary care population with high HIV prevalence in Zimbabwe (see Chibanda et al 2016 DOI: 10.1016/j.jad.2016.03.006).

19. Methods: Please provide a bit more detail on the training of interventionists. Who is conducting the training? Will refresher training be done throughout study follow-up?

The training is conducted by the principal investigators and other clinical psychologists with expertise in the intervention. Refresher trainings and ongoing supervision have been conducted throughout the study.

20. Methods: How is self-reported adherence being assessed?

Self-reported adherence to ART medication at 4-, 8-, and 12-months post randomization is assessed as the frequency of adherence in the past 30 days, measured using a score derived from a three-item questionnaire adapted from Wilson et al. (2015). This has been clarified in the manuscript.

21. Methods: Spell out “SAEs” the first time it is used

This has now been added to the manuscript.

22. Methods – Confidentiality section: Capitalize the “P” in participants (“participants are referred to only by their identification number…”).

This has now been revised in the manuscript

23. Methods – Blinding section: “The IA…and the lead study statistician [are] blinded”

This has now been revised in the manuscript.

24. Methods: What do the authors mean by following an intention to treat principle “as much as possible”?

We thank the reviewer for this comment and have removed the referenced sentence.

25. Methods: Do the authors mean that they will explore moderation by “sex” or by gender?

Moderation will be explored by gender. We have now clarified this in the manuscript.

26. Methods: The authors say “where these data are missing…they will be coded as not suppressed, otherwise the outcome will be left missing”. Can they clarify if all missing data is coded as not suppressed or missing?

We thank the reviewer for this question. If participants have missing data because they have died and their cause of death is AIDS-related we will assume that they were not suppressed. If no data on cause of death is available, or the cause of death is not AIDS related, participants data will be coded as missing.

27. Methods: Please review the line spacing in the “Dissemination” section.

This has now been reviewed.

28. Trial status: Please also state where the trial currently is in terms of enrolment and data collection.

The trial status has now been updated with enrolment data.

Reviewer 3: Dr. Ryan McBain, RAND Corp

1. Based on the authors’ power calculation, they note 85% power to detect an absolute difference of 20% or more in achieving viral suppression. While I understand that pilot data have shown a shift of 25%, this nevertheless strikes me as a very large difference. Are there any other interventions beyond the pilot data that have identified shifts of this magnitude? If so, mentioning these and accompanying citations would be helpful.
The power calculation was based on pilot data, showing a difference of 25%. A difference of 20% of more was chosen as a meaningful difference to present to policy makers.

2. For secondary and tertiary outcomes, do the authors plan to perform a correction for multiple tests? Either way, I would state this in the Statistical Methods section.

We agree with the reviewer that adjustment for multiple testing could be applied to the secondary and tertiary outcomes, though we do not intend to formally do this adjustment as it isn’t currently widely applied in trials.

3. Spacing between lines in the “Dissemination” subsection appears to be different (wider) from the rest of the manuscript.

This has now been reviewed and corrected.

4. The authors mention a preference for reporting mean costs despite (likely) skewed data and cite Thompson and colleagues’ 2000 BMJ article. Could you elaborate more on this justification? (This appears to be an appeal to authority without actually providing the rationale.)

We have included further detail in the manuscript: “Economic evaluations can be used to inform healthcare decision makers on the total budget needed to treat people with a particular disease or condition; it is only the mean cost that allows for this calculation to be made. Thus, it is the arithmetic mean cost that is the relevant summary statistic in pragmatic trials with economic evaluations.”

5. Presumably, the cost-effectiveness component will involve ICERs (incremental cost-effectiveness ratios); however, there is no mention of this. If this is correct, I would suggest adding text to this effect.

The reviewer is correct that the cost-effectiveness analysis would include ICERS – we have amended the text to reflect this.

VERSION 2 – REVIEW

REVIEWER	McBain, Ryan RAND Corp
REVIEW RETURNED	06-May-2022

GENERAL COMMENTS	I am happy to approve of this version of the manuscript with the exception of one detail, which I think is necessary for the authors to address. In my last review, I noted that it is standard practice to correct for multiple comparisons--particularly when examining secondary and tertiary outcomes. Failure to do so would result in high risk of a type-1 error, which I am sure the authors are familiar with. The authors replied to my point by stating: "We do not intend to formally do this adjustment as it isn't currently widely applied in trials." This response seems problematic for at least three reasons. First, I am not convinced that their statement is true. My understanding is that it is common practice to adjust for multiple comparisons, especially if the outcomes are secondary/tertiary. The authorship team has presented no evidence to support their point. Second, even if it were true, replicating an arguably inappropriate practice because others engage in it is not sound rationale; so the authors would need to offer an alternative rationale. Third, even if their response were correct and the rationale were sound, they would need to state their logic in the text to be aboveboard with their readership. In any case, I defer to the editorial team to determine whether the authors' proposed avenue meets muster. As a reviewer, this seems inappropriate to me and don't feel comfortable signing off on this.

VERSION 2 – AUTHOR RESPONSE

Reviewer 3: Dr. Ryan McBain, RAND Corp

1. I am happy to approve of this version of the manuscript with the exception of one detail, which I think is necessary for the authors to address. In my last review, I noted that it is standard practice to correct for multiple comparisons--particularly when examining secondary and tertiary outcomes. Failure to do so would result in high risk of a type-1 error, which I am sure the authors are familiar with. The authors replied to my point by stating: "We do not intend to formally do this adjustment as it isn't currently widely applied in trials."

This response seems problematic for at least three reasons. First, I am not convinced that their statement is true. My understanding is that it is common practice to adjust for multiple comparisons, especially if the outcomes are secondary/tertiary. The authorship team has presented no evidence to support their point. Second, even if it were true, repicating an arguably inappropriate practice because others engage in it is not sound rationale; so the authors would need to offer an alternative rationale. Third, even if their response were correct and the rationale were sound, they would need to state their logic in the text to be aboveboard with their readership.

In any case, I defer to the editorial team to determine whether the authors' proposed avenue meets muster. As a reviewer, this seems inappropriate to me and don't feel comfortable signing off on this.

We thank the reviewer for their close read of the manuscript and our previous response, and agree that a more thorough explanation is needed.

Firstly, regarding tertiary outcomes – these are part of a separate cost-effectiveness analysis, not to be included in the main outcome paper and not focusing predominantly on hypothesis testing/p-values, so we do not think multiple comparisons need to be considered in their case.

Secondly, we note that adjustment for multiple comparisons has been and remains a controversial topic (Schulz & Grimes, 2005; Cramer et al., 2016, Althouse, 2016; Rubin, 2021; Parker & Weir 2022). Generally, these authors point out that while adjustment for multiple testing is necessary in some cases, in others it may not be warranted, and/or may raise other issues. We apologise that we failed before to clearly explain our agreement with Althouse (2016) that the “best approach is simply to (1) describe what was done in a study; (2) report effect sizes, confidence intervals, and p values; and (3) let readers use their own judgment about the relative weight of the conclusions”, and that this was the approach we intended to take.

In addition, we note that our decision of intervention effectiveness will be based on a single pre-specified primary outcome that, Furthermore, we will take the approach outlined by Parker & Weir (2022) in not adjusting for multiple comparisons amongst our secondary outcomes because we will focus on their “precise interpretation” approach, where each outcome is interpreted based on its type, time point, the nature of the intervention, etc. We have added the following to the manuscript

on page 17: “In line with other recent and similar trials in HIV and methodology literature (Schulz and Grimes 2005; Cramer et al. 2016; Althouse 2016; Fox et al. 2018; Saidi et al. 2021; Parker and Weir 2022; Joska et al. 2020), we do not plan to undertake secondary outcome adjustment for multiple comparisons. Rather we will focus on our single pre-specified primary outcome to assess effectiveness and take a “precise interpretation”/separate hypotheses approach for the secondary outcomes in combination with appropriate reporting of effect sizes and confidence intervals to support transparent interpretation (Althouse 2016; Parker and Weir 2022)”.

We have also added several examples of similar trials in HIV (see pages 16 and 17) with multiple secondary outcomes that did not employ adjustment. Two were published in this journal and the examples include both protocol papers and main trial outcome reports.

References included in response to Reviewer 3

(Most also included in the manuscript)

Althouse, A. D. (2016). Adjust for multiple comparisons? It's not that simple. Annals of

Thoracic Surgery, 101(5):1644-1645. doi: 10.1016/j.athoracsur.2015.11.024

Cramer, A. O., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R.

P., Waldorp, L. J., & Wagenmakers, E. J. (2016). Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. Psychonomic Bulletin & Review, 23(2), 640–647. https://doi.org/10.3758/s13423-015-0913-5

Fox, M. P., Pascoe, S. J., Huber, A. N., Murphy, J., Phokojoe, M., Gorgens, M., Rosen, S.,

Wilson, D., Pillay, Y., & Fraser-Hurt, N. (2018). Assessing the impact of the National Department of Health’s national adherence guidelines for chronic diseases in South Africa using routinely collected data: A cluster-randomise evaluation. BMJ Open. Doi: 10.1136/bmjopen-2017-019680.

Joska, J. A., Andersen, L. S., Smith-Alvares, R., Magidson, J., Lee, J. S., O’Cleirigh, C., &

Safren, S. A. (2020). Nurse-delivered cognitive behavioral therapy for adherence and depression among people living with HIV (the Ziphamandla Study): Protocol for a randomized controlled trial. JMIR Research Protocols, 9(2). Doi: 10.2196/14200

Parker, R.A. & Weir, C.J. (2022). Multiple secondary outcome analyses: precise

interpretation is important. Trials, 23, 27. https://doi.org/10.1186/s13063-021-05975-2

Rubin, M. (2021). When to adjust alpha during multiple testing: a consideration of

disjunction, conjunction, and individual testing. Synthese, 6:1–32.

Saidi, F., Mutalie, W., Freeborn, K., Rosenberg, N., Graybill, L. A., Maman, S., Amico, K.

R., Mollan, K. R., Phanga, T., Milala, B., Hill, L. M., Gottwalt, A. M., Phiri, S., Kalua, T., & Chi, B. H. (2021). Combination adherence strategy to support HIV antiretroviral therapy and pre-exposure

prophylaxis adherence during pregnancy and breastfeeding: Protocol for a pair of pilot randomised trials. BMJ Open. Doi: 10.1136/bmjopen-2020-046032.

Schulz, K. F., & Grimes, D. A. (2005). Multiplicity in randomised trials I: Endpoints and

treatments. Lancet, 365, 1591-1595.

VERSION 3 – REVIEW

REVIEWER	McBain, Ryan RAND Corp
REVIEW RETURNED	23-Jul-2022

GENERAL COMMENTS	I think, at this juncture, it would be best to defer to the editorial board on a decision, as I still disagree in principle to the authors' practice of ignoring conventions and not adjusting for multiple corrections when exploring secondary outcomes. Perhaps the editorial board could seek out a statistical reviewer to comment? A couple brief reflections. First, the authors are basing their proposed course of action on an opinion piece in the Annals of Thoracic Surgery written by a non-statistician (Althouse, 2016). A cursory glance at the author's publication record indicates that this is not a topic of his expertise, as none of his other work is directed to this issue. The fact the article is commonly cited, on its face, may merely be evidence that authorship teams have a general incentive to avoid correcting for multiple comparisons when it is commonplace and appropriate to do so. Second, it is my understanding that the significant weight of evidence and ongoing discussion on this topic is dedicated to which and in what ways to correct for multiple comparisons, not whether corrections are appropriate when engaging in exploratory analyses of secondary outcomes. To provide a quote by a senior statistician on the topic publishing recently in a methods journal: "When analysing multiple outcomes, it’s important to control the family wise error rate (FWER)" (Vickerstaff et al, 2019). All said, this seems somewhat like sleight of hand. Put simply, failure to perform any sort of correction for multiple comparisons--especially in the context of secondary, post-hoc analyses--is not appropriate, to the best of my knowledge, and I believe my position is a widely-held view. I appreciate, however, that the authors did take the time to explicitly state a position in their most recent revision.