Author's response to reviews

Title: A combination SMS and transportation reimbursement intervention to improve HIV care following abnormal CD4 test results in rural Uganda: an observational cohort study

Authors:

Mark J Siedner (msiedner@partners.org)
Data Santorino (boymukedata@gmail.com)
Alexander J Lankowski (alexlankowski@gmail.com)
Michael Kanyesigye (mkanyesigye2010@gmail.com)
Mwebesa B Bwana (mwebesa_bwana@yahoo.com)
Jessica E Haberer (jhaberer@partners.org)
David R Bangsberg (dbangsberg@partners.org)

Version: 2  Date: 26 May 2015

Author's response to reviews: see over
Ursula D'Souza, PhD  
Senior Editor  
BMC Medicine

RE: BMC Medicine Manuscript Number 1004543959164530

May 26, 2015

Dear Dr. D’Souza,

On behalf of my co-authors, I thank you for your thorough review of our manuscript. We have responded to each comment in the document which follows.

We look forward to any further comments or requests and appreciate your ongoing consideration of our work.

Sincerely,

Mark J. Siedner  
Instructor of Medicine, Harvard Medical School  
Assistant in Medicine, Massachusetts General Hospital Center for Global Health
Reviewer 1 Comments and Responses (in bold-faced type)

Reviewer's report:
This is an interesting article on a practical yet innovative intervention to improve HIV care in a low-income setting. It is a very good example of implementation science research, highlighting the many challenges and nuances of conducting this type of study. I commend the authors on the thorough mixed methods and collaborative formative research they conducted to design the intervention. There are a number of issues, which, if addressed, would substantially improve the paper. A main point of confusion is whether the primary intent of the study is to make a pre/post comparison OR to compare the 3 different text messaging strategies (i.e. the randomized component). If both are “primary” or perhaps better described as “key” or something to that effect, then I think the abstract and paper should better reflect this. See below for further comments.

We thank the reviewer for their thorough review of our work. We agree with the reviewer that the primary intent of the study was not clear in the prior draft. We have reworded the manuscript (abstract and results sections) to increase the focus on the global effect of the intervention. Details about these changes are discussed in depth below in response to reviewer comments.

Discretionary Revisions
1. Introduction: 3 para, 1 sentence-Wording is off

Thank you for noting this error. We have reworded the sentence to read:

*Scalable interventions that mitigate structural barriers to clinical care in resource-limited settings are urgently needed*

2. Methods: I believe it is spelled Dimagi.

We have corrected that typographic error.

Major Compulsory Revisions (while I would classify these as Major, if the authors have sound reasoning for disagreeing with comments, I would find this acceptable)

1. Title: I found the title a bit misleading. It gives the impression that this is an RCT comparing the intervention to SOC and that two groups are occurring in parallel, when this comparison actually uses a pre/post design. I would suggest deleted “versus Standard of Care” or stating explicitly the main study design.

We have changed the title to respond to this concern. It is now reads:

*A combination SMS and transportation reimbursement intervention to improve HIV care following abnormal CD4 test results in rural Uganda: an observational cohort study*
2. Abstract: A general issue with this study, which is exemplified in the results section in the Abstract, is how the 3 intervention groups are handled, i.e. whether to lump or disaggregate these 3 groups. The current strategy in the abstract is to disaggregate but I think the results section should at least report whether any differences were seen between the 3 randomized groups. It is also not always clear throughout the paper when the authors are reporting a lumped comparison vs. disaggregated.

We agree with the reviewer and have attempted to improve the clarity of the manuscript by increasing focus on the pre-intervention versus intervention period comparisons, and less so on the individual randomized messages, which performed similarly. We have made changes to the abstract and manuscript in multiple locations to reflect this change:

Abstract results section:

There were 45 participants in the pre-intervention period and 138 participants in the post-intervention period with low CD4 count results (46, 49, and 43 in the direct, PIN, and coded groups, respectively). Median time to clinic return was 33 days (IQR 11-49) in the pre-intervention period versus 6 days (IQR 3-16) in the intervention period (P<0.001); and median time to ART initiation was 47 days (IQR 11-75) versus 12 days (IQR 5-19), (P<0.001). In multivariable models, participants in the intervention period had earlier return to clinic (AHR 2.32, 95%CI 1.53 to 3.51) and earlier time to ART initiation (AHR 2.27, 95%CI 1.38 to 3.72). All three randomized message formats improved time to return to clinic and time to ART initiation (P<0.01 for all comparisons versus the pre-intervention period).

Results Section:

The proportion of participants returning with 14 days, the primary outcome of interest, was 27% (12/45) in the pre-intervention period and 67% (93/138) in the intervention period (P<0.001). Median time to clinic return was 33 days (IQR 11 – 49) in the pre-intervention period and 6 days (IQR 3-16) in the intervention period; whereas median time to ART initiation decreased from 47 days (IQR 11-77) to 13 days (5 – 22 days). In multivariable models adjusted for age, gender, district of residence, educational attainment, and CD4 result, participants in the intervention period had shorter time to clinic return (adjusted hazard ratio [AHR]=2.31, 95%CI 1.52 – 3.51, P<0.001, Table 3, Figure 2A) and shorter time to ART initiation (AHR=2.23, 95%CI 1.36 – 3.67, P=0.002, Table 4, Figure 2B).

All three message formats outperformed the pre-intervention period. Median days to clinic return after abnormal results was 4, 11 and 6 days in the direct, PIN and coded message group, respectively (P<0.010 for all comparisons by logrank testing versus the control arm); and median time to ART initiation was 8 (P=0.002), 15 (P=0.016), and 15 days (P<0.001). All analyses remained
statistically significant in multivariable analyses adjusted for sociodemographic and clinical factors (Tables 3A and 3B, Figures 2A-2B). There were no significant differences between message groups in proportion that returned within 14 days (P=0.09), proportion that returned before their scheduled visits (P=0.96), or proportion that initiated ART within 14 days (P=0.32).

We also changed Tables 1 and 2 to focus on the difference between the pre to intervention study periods, and added a Supplementary Table 1 to demonstrate the lack of differences in participant characteristics between the three randomized groups.

3. Methods: Why were clinicians allowed to select the abnormal CD4 results threshold? The reasoning behind the use of this criteria was not clear. It may effect generalizability so further insights into this unusual eligibility criteria would be informative, i.e. what exactly would an “abnormally low result”. Discussing how this might impact on generalizability in the discussion would also be helpful.

The decision to allow clinicians to select the abnormal threshold was made in response to consultation with study clinicians. We held interviews with them prior to the intervention to design a system in response to local needs and improve the likelihood of uptake and use, if effective. They stated a preference for individualized, patient-specific definitions of abnormal test results, largely to allow messaging to patients already on ART, but who were possibly failing therapy to be involved in the program. We have added text to the methods section to better describe the reasoning behind varying the definition of abnormal results:

We discussed message design with clinicians during the intervention design phase. They requested an individualized determination of abnormal test results over a standardized method, largely to allow enrollment of patients already on therapy who were considered to be at risk for treatment failure.

4. In the methods section, it would be helpful to explain the reasoning behind the three SMS options rather than just what they are.

Thank you for this suggestion. We have added text to state the reasoning behind each message type:

Participants were randomized in a 1:1:1 design to receive one of three SMS message formats: 1) an unprotected SMS indicating an abnormal test result and that they should return to clinic as soon as possible: “This is an important message from your doctor. You had an abnormal test result. You should return to clinic as soon as possible” (direct message). This message was meant to maximize clarity; 2) a PIN-protected SMS message, which displayed an identical message as the direct message only after successful entry of the correct PIN code (PIN message). The PIN code was intended to augment message
...privacy; or 3) a message reading “ABCDEFG” (coded message). This message was coded without mention of clinical information to maximize confidentiality, but not require the participant to remember and enter a PIN code.

5. Methods: It would also be good to be clear that your outcome is from test result notification, rather than when the test result was actually performed, e.g. in the Sample Size section one might want to say laboratory result notification rather than just laboratory result.

We agree and have added text to the methods section to clarify this:

To ensure unbiased outcomes assessment between study arms, we changed our primary outcome to: return to clinic within 14 days of when the laboratory result was received back at the clinic from the laboratory.

6. Results: 2nd para-the initial results are presented with the randomized groups disaggregate. I would suggest also providing results with them lumped. Also, as noted above, this results section should preferably report whether there were any differences seen between the 3 intervention groups.

As noted in the response to comment 2 above, we have changed the format of the results section to highlight comparisons between the pre- and intervention periods, and later describe disaggregate comparisons between message groups and the pre-intervention period, as well as within message groups.

7. Results: This paper would benefit from reporting several pieces of process results. For example, do you have data on time from the actual testing to time of return, i.e. was there a significant delay from the time blood was drawn to when the notification went out? Also, and perhaps most importantly, there are no process results on whether text messages were actually received and/or understood. One would like to know, for example, whether the PIN messages were all or mostly unlocked appropriately. As an extreme example, none of the participants might have received any of the text messages and this is actually only a transport reimbursement intervention.

We thank the reviewer for this comment. In response to the first request, we do report time from laboratory test to result return in Table 1 (Days from enrollment until laboratory result), which was a median of 10 days in both study periods. We also agree that more detailed information about message comprehension and acceptance would be informative, but we felt it beyond the scope of this manuscript, which is already quite lengthy. We are currently conducting a post-trial qualitative study with in-depth interviews of study participants to understand barriers and facilitators to the intervention and look forward to presenting that data in a future manuscript.

8. Results: The control group is actually substantially older than the intervention groups and more ART-naïve. Specified threshold is also different. While these may be controlled...
We agree that differences between intervention periods are an important source of bias. However, we do feel that adjustment for these variables in multivariable models is an efficient way of considering this source of bias. We also estimate quite large and statistically significant effect sizes for all study outcomes with adjusted hazard ratios in the 2.0-3.5 range (and corresponding $P$ values < 1/1000), suggesting considerable treatment effects that are unlikely to be caused by residual confounding. That said, we agree with the reviewer and have included text in the limitations section to describe this source of bias:

*Our study is subject to a number of limitations. First, we implemented a non-randomized study design, and found differences in the study periods in participant age and proportion naïve to ART. While the differences in time to clinic return and time to ART initiation could be explained by unmeasured confounders or temporal changes in clinic outcomes, we believe this is unlikely because 1) our results were independent of the most likely confounding variables, including immune status, gender, and education, 2) our effect sizes were large and demonstrated across all randomized sub-groups, and 3) the intervention period immediately followed the control period and no other changes in clinical protocols took place during either period*

9. Results: Table 1- It is not clear from the table what groups are being compared to produce the p value. Suggest adding an additional column which groups the characteristics for all intervention groups together in one column.

We agree and have done so. As described in comment 2, we have changed Table 1 to demonstrate characteristics and comparisons between the pre- and intervention periods only.

10. Results: Table 2-Also, not clear where the p-value is coming from. Would also add a column and lump the intervention groups.

We have changed the format of table 2 to focus on the characteristics in the aggregated pre and post-intervention period and added a footnote to describe the p-values as a comparison between the two.

11. Results: Any data on participant distance from clinic? This would be of interest in regards to generalizability.

We agree with the reviewer. Unfortunately, we did not collect data on distance to clinic. We have previously shown that self-reported distance to clinic is not predictive of clinical outcomes at this site (Siedner et al, AIDS 2013, www.ncbi.nlm.nih.gov/pubmed/23435294) and did not have the funding for home visits to calculate objective measures of distance as part of this study. As a rough
surrogate for distance to clinic, we did include residency within Mbarara district in our multivariable models, and the effect of the intervention was independent of this effect.

12. Discussion: 1st para—again might want to say “abnormal CD4 count notification” if that is what is really going on.

The pre-intervention group was not notified of their results so the term “CD4 count notification” does not apply to many in the study. While the time in our study was calculated from the date of CD4 receipt at the clinic from the central laboratory, this did not differ between study periods. As such, the overall effect is a decreased time to clinic return and ART initiation after implementation of the intervention.

13. Discussion: 2nd para—while this may improve “perceptions of quality of care” you do not have an evidence to this effect. Would modify or delete this statement.

We agree, and have edited the sentence to clarify this was not evaluated in our study but suggested by prior similar work:

> While not evaluated in our study, similar SMS-based notification programs have previously been shown to include secondary benefits, including the potential to improve patient-staff communication and perceptions of quality of care (REFs).

14. Discussion: 2nd para—The fact that those who had normal results also had improved return times may not be because of the SMS messages. They may actually be indicative of a secular change. This should probably be recognized and perhaps included in the limitations section.

We have reworded that sentence to account for this potential bias:

> Preliminary evidence of this effect was suggested by the earlier return times among study participants with abnormal laboratory results, and the higher proportion or participants with normal laboratory who returned within 7 days of their scheduled appointment, although the latter effect could also be the result of other temporal changes.

15. Discussion: 3rd para—I am not sure this is the “first patient-centered, Health intervention demonstrated to improve patient-provider clinical communication”. Did the WelTel study not do this?

We agree and have reworded the sentence to state it is “among the first” as opposed to the first.

16. In the study limitations, I would probably also discuss the risk of bias given the study was unblinded.
We agree with the reviewer and have specified in the limitations that the study was unblinded:

First, we implemented an unblinded, non-randomized study design, and found differences in the study periods in participant age and proportion naïve to ART.

17. Trial Registration-Finally, I will note that this study is registered in Clinicaltrials.gov which is good. However, some of the data from clinicaltrials.gov does not exactly match up with the paper. For example, the primary outcome on the website says receipt and comprehension of text-related message (which do not appear to be reported in this manuscript) rather than time to return to clinic and time to art initiation (outcomes which I could not find on the website).

When the study was initially designed, it was designed with two primary goals in mind: 1) to evaluate the acceptability and feasibility of the intervention, which includes message receipt and comprehension which we intend to report elsewhere; and 2) to estimate the efficacy of the intervention, which we are reporting in the current manuscript. The primary efficacy outcome was time to clinic return, which we report in this manuscript. While we did not initially expect to be powered to detect a change in time to ART initiation during study design, we were admittedly surprised by the effect identified, and felt it was in important finding to include with our results.

Quality of written English: Acceptable
Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.
Reviewer's report:
This is a quasi experimental study of the use of mobile phones + transportation incentives to promote retention among HIV patients in Uganda. The paper is well written, I enjoyed reading it, I think it has important implications for the health system and public health.

We thank the reviewer for her review and support of our work.

Major revision:
1) What was the rationale of having three different subgroups for the intervention? One sub group in particular does particularly better than the other two. Why? There was no discussion of this.

We thank the reviewer for raising this important point. An important secondary goal of the study was to respond to feedback from study participants about message privacy raised during pre-intervention interviews. As such we included three separate message formats which varied in aspects of privacy and confidentiality. We have added information to the manuscript to better explain this reasoning:

Pre-intervention interviews with patients and providers at the clinic revealed two major barriers to clinic return after abnormal test results at this site: 1) lack of communication between staff and patients outside of clinical visits, which limited patient decision-making capacity, and 2) inadequate access to financial support for costs of transportation to clinic… After the basic features of the intervention were selected, we implemented a co-creation model involving researchers, programmers, clinicians and patients in key considerations about the SMS intervention, including 1) optimizing message formats to maximize confidentiality and privacy, 2) selection of the frequency, duration, and timing of messages, 3) language preferences, and 4) including an option to respond to messages for confirmation of receipt. We pilot tested the system with both study staff and clinicians to optimize interpretation, readability, and ease of use.

Participants were randomized in a 1:1:1 design to receive one of three SMS message formats: 1) an unprotected SMS indicating an abnormal test result and that they should return to clinic as soon as possible: “This is an important message from your doctor. You had an abnormal test result. You should return to clinic as soon as possible” (direct message). This message was meant to maximize clarity; 2) a PIN-protected SMS message, which displayed an identical message as the direct message only after successful entry of the correct PIN code (PIN message). The PIN code was intended to augment message privacy; or 3) a message reading “ABCDEFG” (coded message). This message was coded without mention of clinical information to maximize confidentiality, but not require the participant to remember and enter a PIN code.

We have attempted to clarify the results section by highlighting both the differences
between the pre- and intervention period, but also to better demonstrate differences between and within sub-groups. Notably there were no statistical differences in primary outcomes between the different message groups and we have added text to this effect in the manuscript results section:

*There were no significant differences between message groups in proportion that returned within 14 days (P=0.09), proportion that returned before their scheduled visits (P =0.96), or proportion that initiated ART within 14 days (P=0.32).*

2) There is no mention of attrition in any of the groups, the control or any of the intervention ones. Does this mean that everyone actually did come back, even in the control group? This would imply that patients are not being lost but rather are coming back later than they should. Does coming back 2-3 weeks later make a clinically significant difference? What are attrition rates normally at the center?

There was very little loss to follow-up in the study. We did re-allocate any participant who was tracked at home at 28 days as a failure and have included this information in the manuscript:

*The number of participants who were tracked at home at study day 28 (and allocated as failures to initiate ART) was 2 (4%), 3 (6%), and 1 (2%) in the direct, PIN, and coded arms*

While we agree with the reviewer that our study did not demonstrate changes in loss to follow-up, we do feel that the impact of the intervention had meaningful clinical outcomes. For example, the intervention decreased return to clinic after abnormal CD4 results from a median of 33 to 6 days and time to ART initiation time from a median of 47 to 11 days. Particularly in patients with low CD4 counts not yet on ART, in whom rates of opportunistic infection range up to 100/100 person-years (see: Mocroft et al, Lancet, 2000, www.ncbi.nlm.nih.gov/pubmed/11071184), decreasing time to ART initiation by a matter of weeks can be quite meaningful.

Minor comments:
1) Could the fact that a larger proportion of the control group were already on ART have influenced the delay to return to the clinic compared to the other groups? Perhaps this group was more complacent because of the 'sense of security' provided by being on meds.

We agree with the reviewer about this concern. We accounted for this in two ways. First, we included the prior CD4 count and whether the participants had previously received ART in our multivariable models. While being ART naïve did improve time to return to clinic (AHR 1.48, 95% CI 1.00 – 2.19) the effect of the intervention was independent of this in multivariable models (AHR 2.32, 95% CI 1.53 – 3.51).

Second, our time to ART initiation outcome excluded those with prior ART use. In these models, the effect of the intervention remained strongly significant (AHR 2.27, 95% CI 1.38 – 3.72). See Tables 3A-3B for these results.
2) What could have been reasons for difference between the control and intervention groups besides the intervention. The discussion is 'local' at the moment, I think it would benefit from a widening of scope to talk about the implications for other settings. Have their been other mhealth interventions tested for retention? Without the transport support component? Was the effect size detected as large?

This is an important point. First, while our study was not randomized due to logistical and safety concerns of randomizing participants to not receiving messages (as described in the methods section), we feel that the effect size makes it reasonably unlikely that other factors can explain the effect:

While the differences in time to clinic return and time to ART initiation could be explained by unmeasured confounders or temporal changes in clinic outcomes, we believe this is highly unlikely because 1) our results were independent of the most likely confounding variables, including immune status, gender, and education, 2) our effect sizes were large and demonstrated across all randomized sub-groups, and 3) the intervention period immediately followed the control period and no other changes in clinical protocols took place during either period.

We are not aware of other SMS-based interventions specifically designed to address retention in HIV care (as opposed to adherence). As described in the methods section, we developed the combination intervention in response to participant interviews, which noted both communication and structural barriers to return. As such, we have acknowledged that it is impossible to separate the two elements of our combined intervention, the SMS message and the transportation stipend, and that future studies should assess for an independent effect of the SMS program (which we are planning).

We developed and tested a combination intervention, so cannot assess whether the impact we detected was a result of either intervention alone or the combination of them. However, we developed the combination intervention in response to a conceptual framework derived from formative research noting both financial and communication barriers to care, and thus posited both were needed to meaningfully impact care. For example, a transportation reimbursement for early clinic return would be of little benefit in the absence of a notification system to alert patients of the need for early return. An important area of future study will be to identify the independent effects of the SMS messaging component with and without a transportation stipend.

Quality of written English: Acceptable
Statistical review: Yes, and I have assessed the statistics in my report.