PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Management of type 2 diabetes in China: The Happy Life ClubTM, a pragmatic cluster randomised controlled trial using health coaches
AUTHORS	Browning, Colette; Chapman, Anna; Yang, Hui; Liu, Shuo; Zhang, Tuohong; Enticott, Joanne; Thomas, Shane

VERSION 1 - REVIEW

REVIEWER	Peter Rohloff Brigham and Women's Hospital, USA Maya Health Alliance, Guatemala
REVIEW RETURNED	15-Aug-2015

GENERAL COMMENTS	I enjoyed reading this interesting paper on an important topic, interventions to improve self efficacy and diabetes control in China. This is a topic that is relevant to all of us working in the emerging epidemic of noncommunicable diseases in transitioning economies and evolving health care systems. The study design, a practice-level cluster randomized controlled trial, is excellent for the research topic and the implementation environment. However, many of the trial design assumptions were violated in the implementation (due to factors beyond control of the authors), I have the following queries and comments for the authors: - This was a negative trial, based on failing to achieve statistically significant outcomes in the primary endpoint (A1C) - I'd suggest making this clearer in the abstract and discussion. - Table 2 is very "busy" - I wonder if the authors might consider reformatting this to include less data ,and include the full data set as an appendix or supplementary information? For example, since the abstract and discussion focus on changes at 12 months, it might be possible to just report the baseline measurements and the mean change from baseline to 12 months (+/- CI) and leave the rest for supplementary information. This would definitely make readers' perusal of the salient details easier. - In the study protocol (BMC Public Health 11:90; 2011) there was mention of other metrics, like self-efficacy, disease related knowledge, self care activities. Were these assessed? If so it would be really interesting to have those here. This is because the reported differences between the two groups are fairly trivial in terms of secondary outcomes and it would be interesting to see if the MI produced other knowledge or self-care related effects that might

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

	have a longer time-course in terms of seeing a difference. - As it stands the trial seems primarily to have served as a catalyst for reconstituting diabetes primary care. The very high mean A1C at baseline (>10) in both groups suggests that subjects in neither group were receiving any diabetes care at all. The dramatic reductions in the primary outcome in both control and intervention care is consistent with access to primary care and not much else. This is a major and salutary achievement, but it does make interpretation of the impact of MI almost impossible. - As a global health practitioner myself, what I'd love to see throughout the article is more explicit discussion of how the trial itself influenced patient recruitment or the dynamics of clinical practice. This seems much more important to me than small changes in secondary outcomes between the groups. These insights could help those of us who practice in global health contexts design similar interventions elsewhere. I don't mean "how the trial" influenced in a pejorative sense, because the clinical impact of the trial for both arms is very impressive. But rather in a salutary sense, of how research programs can revitalize primary care in emerging health care contexts. This seems like a major take away point. - I'm not familiar with the significance of the Kessler scale. Can the authors comment on what (if any) clinical significance a +0.8 vs +2.56 might indicate? - Can the authors comment on the significant difference in profile between their pilot and this larger trial? Again I'm interested in this from an implementation science perspective, and learning things from this trial that I can apply in my own setting. Why is the mean A1C in the published pilot (Front Public Health. 2014; 2: 181) so much better ("more controlled") than the mean A1C in this larger trial? One could hypothesize for example that positive media attention to the trial meant that the PHCs were flooded with an influx of new, untreated patients, and new higher patient volume which essential changed all of the underlying assumptions that formed the basis for the trial. This could be looked at by giving us some info on PHC patient flow (number of new patients, that sort of thing) if available. Again, from a pragmatic point of view, this is super important because these are the kinds of things that happen in the real world. If the authors could be more explicit about what happened that would be of tremendous value to all of us real world implementers. - I know the authors say that medication use was not assessed, but were any other dosage variables assessed - for example mean number of doctor visits per year at baseline and then during the trial? This gets to my earlier point that this is really a trial of reconstituting primary care (with a marginal if any effect of motivational intervention over and above just getting good primary care). If this kind of data is available, I think the paper would be much stronger by shifting the tone towards this insight as the primary one in the discussion: "Access to primary care was significantly overestimated in trial design and so, when it was reconstituted in a pragmatic trial the resulting inherent major changes in the primary outcome, overwhelmed the statistical assumptions of the trial design preventing robust analysis of the impact of motivational interviewing."

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

	- Was any auditing of the "MI" dose and practice characteristics and/or protocol fidelity of providers and PHCs conducted? Again this gets to the implementation science piece - is this trial just showing that, when monitored and supported providers overall do a better job, independent of whether they do MI? Where there significant differences in provider-patient interactions in the MI arm? - I'm a clinician, not a statistician, so this comment may be totally ignorant: would some post-hoc analysis of the sample size/power calculations be useful? Given that many of the central assumptions of the sample size calculation were invalidated (major off target effect of simply being in a trial. ICC for A1C at baseline of 0.35 (!), large variation in size of clusters). - Can the authors provide us with some insights into how effective MI trials in settings like this might be done? Again, I'm not a statistician, but as a clinician I would find this discussion helpful- feel free to ignore if the comment is uninformed! How can we effectively assess the impact of MI in the absence of primary care systems? Are there alternative statistical designs that might help us adequately power trials when we anticipate large changes from baseline in both control and intervention groups. Is MI even necessary/helpful prior to just providing patients with access to doctors and medications? For example, a statistical design which looks at differences in rates of decline in A1C (to account for a large off target effect size of the intervention/control) I suspect would have a much larger sample size requirement, large enough that it might make this kind of a trial financially or logistically unfeasible if adequately powered for the primary outcome. - Can the authors update us on how things are going now, or what continuation plans might be? I'm interested in the long-term impacts of this trial. Are PHCs still providing better/enhanced care? Have patients all be lost to followup, or are they still in care? After the media rush, did PHCs see their patient volume subside again to the pre-trial baseline. Again, these would be important insights from a longer term implementation science perspective as we thing about the impact of research activities on local health systems (in terms of capacity builidng, and whether that is sustained or no when the trial ends). -Figure 3B and 3C are misleading without confidence intervals on the outcomes, because most of these outcomes differences are not significant and some readers may jump right to the figures without reading the text. Thanks so much for this important work!

REVIEWER	Lise Juul Aarhus University, Department of Public Health, Denmark
REVIEW RETURNED	02-Oct-2015

GENERAL COMMENTS	The authors aim to assess the effectiveness of a coach-led motivational interviewing intervention in improving glycaemic control, as well as clinical and psychological outcomes of patients with type 2 diabetes compared with usual care. The study addresses an

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

	important subject and adds some interesting findings. The MI-intervention showed no effect on the primary outcome HbA1c, but positive effects on psychological distress and systolic blood pressure. Very interestingly, an overall very large mean reduction in HbA1c in a very short time was shown. I think that the key findings could be more highlighted and discussed; for example the challenges of performing pragmatic trials and the possible consequences on the results. The authors´ definition of the “pragmatic design” could in advance be more explicit. If the authors aimed to assess whether the intervention worked in “real life”, the sentence page 4, line 12; “The heterogeneous population sampled in this pragmatic trial, as well as the potential variability of intervention delivery between CHSs may have diluted the observed treatment effects” makes little sense. Is the definition of “pragmatic design” in accordance with the paper “Loudon K., et al. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015;350:h2147” ? The design choices regarding recruitment of Community Health Stations (CHS) and participants are very much in line with the pragmatic approach. However, a flowchart of the participant recruitment is missing. How many of the eligible participants were contacted, declined etc. Was it a heterogeneous study population? I think a discussion on how representative the included population ended up being is missing (despite the good intentions). Furthermore, I think that the incentive that participation in the trial required no payment and payment is usually required for CHS visits in China (described in the protocol paper) needs even more attention in the paper (and in the abstract). Also the major reductions in HbA1c need more attention and discussion. May these very large reductions in HbA1b in both groups be explained by a very motivated study population or/and by “the intervention point” of free regular consultations? (and is it without harm?). More clear descriptions of the intervention and the comparison practice, highlighting both “the intervention” in both groups and the difference between the groups could be provided (maybe graphical; for inspiration see “Perera R, Heneghan C, Yudkin P: Graphical method for depicting randomised trials of complex interventions. BMJ 2007;334:127-129”). The quite intense follow up/monitoring is mentioned but not discussed according to the pragmatic design. Neither are the choice and the reporting of the secondary outcomes. It seems that there are more collected outcomes than reported? For example if collected as planned according to the Trial Registration, outcomes such as Diabetes management self-efficacy, Diabetes self care activities and Lifestyle factors could have been used to investigate the expected mechanisms, the underlying rationale of the intervention. Why was the measure on psychological distress chosen to be reported? Why not self-rated health and quality of life? A description of “The Kessler 10 “ is missing. Also, discussions on the clinical relevance of the 2.66-point between-group difference in this outcome (it may be a relevant difference in a pragmatic design?) and the possible consequence of non-blinding on the result of this self-reported outcome are missing. Likewise, the possible consequence of non-blinding of data collectors on the blood pressure result could be discussed. Finally, maybe, the slightly larger percentage (6%-points) of missing data in the intervention group compared to the control also is relevant to discuss?

REVIEWER	Chanand K. Saha

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

	Indiana University School of Medicine USA
REVIEW RETURNED	09-Oct-2015

GENERAL COMMENTS	This is an excellent study including very sound study design, sample size justification and appropriate statistical analyses. The study was very carefully designed to ensure the intervention and control groups are very similar to each other. Statistical methods took into account the correlation at two levels, clinic level and repeated measurements at the individual level. The study did not show any differential treatment effect in primary outcome, improvement in HbA1c, though both groups showed very impressive improvement both at 6- and 12-month. In addition, none of the secondary clinical outcomes showed any differential treatment effect except for SBP, but due to very small magnitude of difference and looking at a large number of secondary outcomes, this finding might be attributed to large type I error. The authors claimed that they observed a significant improvement in intervention subjects compared to control subjects in psychological distress outcome, but none of the two groups showed any improvement in psychological distress, table 2 shows no significant change at 6-month and a small magnitude of worsening score for the intervention group and a large magnitude of worsening score for the control group at 12-month. Though the authors claimed there might be a dilution treatment effect due to possible contamination or drawing similar attention to the control subjects, it should not be concluded that intervention was better than the usual care. While there was very impressive improvement in HbA1c for both groups, there was either no or very minimal improvement in fasting glucose. Can the authors provide any explanation for such observation?

VERSION 1 – AUTHOR RESPONSE

Reviewer One

• This was a negative trial, based on failing to achieve statistically significant outcomes in the primary endpoint (A1C) - I'd suggest making this clearer in the abstract and discussion.

To address this comment, the abstract and main text (both results and discussion) have been restructured to report on the primary outcome first (of which was negative), followed by the secondary outcomes.

As an example, the results section of the abstract (page 3) now reads:
"At 12 months, no differential treatment effect was found for HbA1c (adjusted difference -0·11, 95% CI -0·94 to 0·71, p=0·79), with both treatment and control groups showing significant improvements. However, two secondary outcomes: psychological distress (adjusted difference -2·66, 95% CI -4·97 to -0·35, p=0.02) and systolic BP (adjusted difference -3·37, 95% CI -5·84 to -0·90, p=0.01) were robust outcomes consistent with significant differential treatment effects, as supported in sensitivity analyses. Interestingly, in addition to HbA1c, both groups displayed significant improvements in triglycerides, LDL cholesterol and HDL cholesterol."

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

For changes to the results and discussion, please refer to the revised article (results: from beginning of page 16, and first paragraph of discussion).

• Table 2 is very "busy" - I wonder if the authors might consider reformatting this to include less data, and include the full data set as an appendix or supplementary information? For example, since the abstract and discussion focus on changes at 12 months, it might be possible to just report the baseline measurements and the mean change from baseline to 12 months (+/- CI) and leave the rest for supplementary information. This would definitely make readers' perusal of the salient details easier.

Table 2 (clinical data) and Table 3 (psychosocial data) have been reformatted as per this suggestion. Six month data (and mean change from baseline-six months) have been removed from the table in the main text and have been included as a supplementary table.

• In the study protocol (BMC Public Health 11:90; 2011) there was mention of other metrics, like self-efficacy, disease related knowledge, self-care activities. Were these assessed? If so it would be really interesting to have those here. This is because the reported differences between the two groups are fairly trivial in terms of secondary outcomes and it would be interesting to see if the MI produced other knowledge or self-care related effects that might have a longer time-course in terms of seeing a difference.

These additional measures have now been added to the paper. Table 2 now presents the full suite of clinical (physical) data and Table 3 presents the full suite of data from the psychosocial measures.

• As it stands the trial seems primarily to have served as a catalyst for reconstituting diabetes primary care. The very high mean A1C at baseline (>10) in both groups suggests that subjects in neither group were receiving any diabetes care at all. The dramatic reductions in the primary outcome in both control and intervention care is consistent with access to primary care and not much else. This is a major and salutary achievement, but it does make interpretation of the impact of MI almost impossible.

This point has been added to the discussion (page 27, para 1):
"The mean HbA1c at baseline of >10% in both groups suggests that participants in our sample were either not accessing CHSs for T2DM management or were receiving suboptimal T2DM care prior to enrolment in this trial. Accurate health service utilisation records were not able to be obtained for the period before our study; however it can be assumed that this trial served as a catalyst for the revitalisation of primary care delivery to individuals with T2DM. While the possible revitalisation may be attributed to a multitude of practice, participant, and study-related factors, the interpretation of the true effect of MI is consequently limited."

• As a global health practitioner myself, what I'd love to see throughout the article is more explicit discussion of how the trial itself influenced patient recruitment or the dynamics of clinical practice. This seems much more important to me than small changes in secondary outcomes between the groups. These insights could help those of us who practice in global health contexts design similar interventions elsewhere. I don't mean "how the trial" influenced in a pejorative sense, because the clinical impact of the trial for both arms is very impressive. But rather in a salutary sense, of how research programs can revitalize primary care in emerging health care contexts. This seems like a major take away point.

Revitalising primary care was most likely a sequela to the research program (as described in the previous bullet point). In a way, Reviewer 2 (bullet point 5) also brought up this issue. Also Reviewer 3

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

(bullet point 3). This question will be fully addressed in a subsequent qualitative paper that will investigate how the trial influenced (and continues to influence) clinical practice - from the perspectives of both clinician and patient.

• I'm not familiar with the significance of the Kessler scale. Can the authors comment on what (if any) clinical significance a +0.8 vs +2.56 might indicate?

The score range for Kessler 10 (and the additional psychosocial measures) has now been added to the methods section (page 10, bottom paragraph 1). A full description of the psychosocial measures have also been added as a supplementary document (see supplementary file).

Additionally, the clinical significance of the significant increase in psychological distress in the control group has been discussed as follows (discussion paragraph 2, bottom of page 25, top of page 26):
"In addition to being statistically significant, greater psychological distress observed in the control group compared to the intervention group is of clinical significance. The shift in the mean score for the control group at baseline from 14·97 ± 6·24 to 17·45 ± 8·12 at 12 months translates clinically to a shift in the mean from "low risk of psychological distress" (scores 10-15) to "moderate risk of psychological distress" (scores 16-21)."

• Can the authors comment on the significant difference in profile between their pilot and this larger trial? Again I'm interested in this from an implementation science perspective, and learning things from this trial that I can apply in my own setting. Why is the mean A1C in the published pilot (Front Public Health. 2014; 2: 181) so much better ("more controlled") than the mean A1C in this larger trial? One could hypothesize for example that positive media attention to the trial meant that the PHCs were flooded with an influx of new, untreated patients, and new higher patient volume which essential changed all of the underlying assumptions that formed the basis for the trial. This could be looked at by giving us some info on PHC patient flow (number of new patients, that sort of thing) if available. Again, from a pragmatic point of view, this is super important because these are the kinds of things that happen in the real world. If the authors could be more explicit about what happened that would be of tremendous value to all of us real world implementers.

Thank you for raising this. We have included a paragraph in the discussion (page 29, para 2) that addresses this point.
"Although our pilot study utilised the same pragmatic design as the present trial, some notable differences in the results for HbA1c were observed when comparing trials. In the pilot study, a significant improvement in HbA1c from baseline to 6 months was observed in the intervention group only. Additionally, the baseline HbA1c among both treatment groups in the pilot study was substantially lower (~7.0%) than that observed in the present trial (~10%). One explanation for this variation may be the difference in the quality of care delivered by the CHSs in the pilot and the present trial. The pilot study was conducted in the Fangzhuang Community Health Centre, which was the only nationally certified centre in the Fengtai district at the time of both studies.33 As such, it is possible that the usual care provided in the pilot to individuals with T2DM was of higher quality than that offered by the CHSs in the present trial. Another possibility is that in the pilot compared to the full trial less "leakage" occurred across the treatment and control arms of the study. The study was widely reported across the Chinese media and this may have impacted upon treatment fidelity in the control arm of the full study."

• I know the authors say that medication use was not assessed, but were any other dosage variables assessed - for example mean number of doctor visits per year at baseline and then during the trial? This gets to my earlier point that this is really a trial of reconstituting primary care (with a marginal if any effect of motivational intervention over and above just getting good primary care). If this kind of data is available, I think the paper would be much stronger by shifting the tone towards this insight as

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

the primary one in the discussion: “Access to primary care was significantly overestimated in trial design and so, when it was reconstituted in a pragmatic trial the resulting inherent major changes in the primary outcome, overwhelmed the statistical assumptions of the trial design preventing robust analysis of the impact of motivational interviewing.”

We agree that this would have made a useful addition to the paper. Unfortunately, accurate data is not available for any other dosage variables such as health service utilisation. The unavailability of this data has been stated in the discussion (page 27, para 1): “Accurate health service utilisation records were not able to be obtained…”

• Was any auditing of the “MI” dose and practice characteristics and/or protocol fidelity of providers and PHCs conducted? Again this gets to the implementation science piece - is this trial just showing that, when monitored and supported providers overall do a better job, independent of whether they do MI? Where there significant differences in provider-patient interactions in the MI arm?

We agree that this is an important aspect to consider. All MI sessions (both face-to-face and telephone) were audio-recorded and an assessment of treatment integrity is planned in the near future and will be reported on in a subsequent publication. This assessment will utilise the Motivational Interviewing Treatment Integrity (MITI) Framework. This will be a large piece of work, and when completed, a separate paper will be prepared and submitted. This future planned piece of work has been listed as a future piece of work in the discussion section (page 31, para 2)

• I’m a clinician, not a statistician, so this comment may be totally ignorant: would some post-hoc analysis of the sample size/power calculations be useful? Given that many of the central assumptions of the sample size calculation were invalidated (major off target effect of simply being in a trial. ICC for A1C at baseline of 0.35 (!), large variation in size of clusters).

This is an interesting idea that we consulted with our study statistician about. The advice was that when we look at the magnitude of the differences between both the treatment and control arms, the differences are very small and in most case negligible. Increasing the sample size would likely not alter these effect sizes. It is well known that in very big sample trials, significant differences are the mainstay even when effect sizes are incredibly small. In these large trials, as with all trials really, we need to consider if effect sizes are meaningful (i.e. small effect sizes may be statistically significant but not clinically meaningful). Also, in this pragmatic trial, increasing the sample size will not overcome the main sources of variance which are likely from non-constant treatment fidelity and other factors outside of our control (see response to comment immediately below as this is somewhat related).

• Can the authors provide us with some insights into how effective MI trials in settings like this might be done? Again, I’m not a statistician, but as a clinician I would find this discussion helpful- feel free to ignore if the comment is uninformed! How can we effectively assess the impact of MI in the absence of primary care systems? Are there alternative statistical designs that might help us adequately power trials when we anticipate large changes from baseline in both control and intervention groups? Is MI even necessary/helpful prior to just providing patients with access to doctors and medications? For example, a statistical design which looks at differences in rates of decline in A1C (to account for a large off target effect size of the intervention/control) I suspect would have a much larger sample size requirement, large enough that it might make this kind of a trial financially or logistically unfeasible if adequately powered for the primary outcome.

Study designs that utilise an experimental RCT design (where standard operating systems are tightly controlled), in preference to a pragmatic RCT, are more likely to observe larger changes favouring the intervention. While pragmatic trials are designed to assist in supporting a decision on whether to

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

deliver the chosen intervention as part of routine care, explanatory trials are designed to test causal research hypotheses under ideal conditions. By selecting to utilise a pragmatic trial design, our intention was to determine if our chosen intervention would work in the setting for which it was to be applied. We have added a justification to our methods section (page 7, para 1), to more clearly state our purpose of selecting a pragmatic design.

"The selection of a pragmatic trial design, which is undertaken in the 'real world' and with usual care,17 also suited the context of the intervention site, namely Community Health Stations (CHSs) within a district of Beijing where preventive care, health management, primary medical care, rehabilitation, health education, and family planning are offered.18 Additionally, the intention of this pragmatic study was to assist in supporting a decision on whether to deliver the chosen intervention as part of routine care."

• Can the authors update us on how things are going now, or what continuation plans might be? I'm interested in the long-term impacts of this trial. Are PHCs still providing better/enhanced care? Have patients all be lost to follow-up, or are they still in care? After the media rush, did PHCs see their patient volume subside again to the pre-trial baseline? Again, these would be important insights from a longer term implementation science perspective as we thing about the impact of research activities on local health systems (in terms of capacity building, and whether that is sustained or no when the trial ends).

Any future investigations among the CHSs in the present trial will be dependent on the attainment of further funding. However, an extension of the present trial will be implemented in Shenzhen in 2016/17 as part of the functions of the newly formed International Primary Health Care Research Institute. This point has been added to the concluding sentence of the paper (page 32).

• Figure 3B and 3C are misleading without confidence intervals on the outcomes, because most of these outcomes differences are not significant and some readers may jump right to the figures without reading the text.

Confidence intervals have now been added to these figures.

Reviewer Two

• I think that the key findings could be more highlighted and discussed; for example the challenges of performing pragmatic trials and the possible consequences on the results.

The abstract and main text (both results and discussion) have been restructured to report on the primary outcome first (of which was negative), followed by the secondary outcomes. Additionally, both the methods and the discussion have been updated throughout to include more mention of the pragmatic design of this trial.

• The authors´ definition of the "pragmatic design" could in advance be more explicit. If the authors aimed to assess whether the intervention worked in "real life", the sentence page 4, line 12; "The heterogeneous population sampled in this pragmatic trial, as well as the potential variability of intervention delivery between CHSs may have diluted the observed treatment effects" makes little sense. Is the definition of "pragmatic design" in accordance with the paper "Loudon K., et al. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015;350:h2147" ?

The definition of the 'pragmatic design' has been added to the first paragraph of the methods (page 7) following the justification of a cluster design. The pragmatic design utilised in this study was originally guided by the CONSORT extension for pragmatic trials and the original PRECIS tool published in

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

2009. However, the definition of the pragmatic design is inline with the more recent publication you suggested above and as such we have cited this publication when defining the pragmatic design. Further, the statement on page 4 has also been removed.

• The design choices regarding recruitment of Community Health Stations (CHS) and participants are very much in line with the pragmatic approach. However, a flowchart of the participant recruitment is missing. How many of the eligible participants were contacted, declined etc. Was it a heterogeneous study population? I think a discussion on how representative the included population ended up being is missing (despite the good intentions).

We agree that this would be a useful addition to the paper according to the pragmatic nature of the trial. However, the majority of CHSs unfortunately had incomplete recruitment records regarding the number of people contacted for participation. As a result, we are unable to accurately prepare a flow chart of participant recruitment. The following statement has been added to the results section (page 14, results - para 2) to address this point:
“In line with the recommendations for the reporting of pragmatic trials, numbers of eligible patients and numbers of those who were contacted and declined (along with the reasons for non-participation) were intended to be reported on; however incomplete recruitment records in a considerable number of CHSs prevented this from occurring.”

• Furthermore, I think that the incentive that participation in the trial required no payment and payment is usually required for CHS visits in China (described in the protocol paper) needs even more attention in the paper (and in the abstract).

We agree that this would be a useful addition and have updated both the abstract and the main text accordingly.
In the abstract, the following sentence has been added:
“Medical fees were waived for both groups.”

Within the main text the following sentence has been added (page 13, para 2):
“Although China has near universal health insurance coverage, individuals with T2DM typically incur out-of-pocket expenses for both medical and pharmaceutical care.”

• Also the major reductions in HbA1c need more attention and discussion. May these very large reductions in HbA1c in both groups be explained by a very motivated study population or/and by “the intervention point” of free regular consultations? (and is it without harm?).

The results and discussion have been updated to address the primary outcome (HbA1c) first, followed by the secondary outcomes. Additionally, throughout the results and discussion, more specific mention and discussion of HbA1c has been made. While attention to HbA1c has been heightened throughout the paper, an example of this is (page 27, para 1):

“Although glycaemic control did not differentially improve, HbA1c in both groups changed significantly and for the better, as did triglycerides, LDL cholesterol and HDL cholesterol. The mean HbA1c at baseline of >10% in both groups suggests that participants in our sample were either not accessing CHSs for T2DM management or were receiving suboptimal T2DM care prior to enrolment in this study. Accurate health service utilisation records were not able to be obtained for the period before our study; however it can be assumed that this study served as a catalyst for the revitalisation of primary care delivery to individuals with T2DM. While the possible revitalisation may be attributed to a multitude of practice, participant, and study-related factors, the interpretation of the true effect of MI is consequently limited.”

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

• More clear descriptions of the intervention and the comparison practice, highlighting both “the intervention” in both groups and the difference between the groups could be provided (maybe graphical; for inspiration see “Perera R, Heneghan C, Yudkin P: Graphical method for depicting randomised trials of complex interventions. BMJ 2007;334:127-129”).

Thank you for suggestion and the useful reference. As per your recommended article, Figure 2 has been updated and now graphically depicts the intervention conditions for both treatment groups.

• The quite intense follow up/monitoring is mentioned but not discussed according to the pragmatic design.

More reference has now been made throughout the paper with regard to the pragmatic design, including the intense follow up monitoring (of which deviates from standard usual care).

• Neither are the choice and the reporting of the secondary outcomes. It seems that there are more collected outcomes than reported? For example if collected as planned according to the Trial Registration, outcomes such as Diabetes management self-efficacy, Diabetes self-care activities and Lifestyle factors could have been used to investigate the expected mechanisms, the underlying rationale of the intervention. Why was the measure on psychological distress chosen to be reported? Why not self-rated health and quality of life?

The additional psychosocial outcomes have now been added to the paper. Table 2 presents the full suite of clinical data and Table 3 the full suite of psychosocial data.

• A description of “The Kessler 10 “is missing. Also, discussions on the clinical relevance of the 2.66-point between-group difference in this outcome (it may be a relevant difference in a pragmatic design?)

The score range for Kessler 10 (and the additional psychosocial measures) has now been added to the methods section (page 10, bottom paragraph 1). A full description of the psychosocial measures have also been added as a supplementary document (see supplementary file).

Additionally, the clinical significance of the significant increase in psychological distress in the control group has been discussed (discussion paragraph 2, page 25/26):
“In addition to being statistically significant, greater psychological distress observed in the control group compared to the intervention group is of clinical significance. The shift in the mean score for the control group at baseline from 14·97 ± 6·24 to 17·45 ± 8·12 at 12 months translates clinically to a shift in the mean from “low risk of psychological distress” (scores 10-15) to “moderate risk of psychological distress” (scores 16-21).”

• The possible consequence of non-blinding on the result of this (K10) self-reported outcome are missing. Likewise, the possible consequence of non-blinding of data collectors on the blood pressure result could be discussed.

The following statement has been added to the limitations in the discussion (page 31, para 2):
“Further, the lack of blinding of outcome assessors as well as the limitations inherent in self-reported data are worth noting when interpreting the findings.”

• Finally, maybe, the slightly larger percentage (6%-points) of missing data in the intervention group compared to the control also is relevant to discuss?

A missing data analysis was done and showed that dropouts were missing completely at random

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

despite group allocation (Little's MCAR test: Chi-Square = 19·4, df = 18, p = 0·4). This result means that the results obtained (even with approx. 6% more missing data in the control group) appeared to be representative. The missing data analysis is reported in the results section (2nd paragraph), and is now also mentioned in the methods section at the very end:

"A missing data analysis was done for each outcome measure and consisted of Little's MCAR test to investigate patterns of missingness in variables of group allocation (intervention/control) and baseline characteristics as listed in Table 1. Sensitivity analyses were done using multiple imputation to account for missing data and then re-running the analyses….."

Reviewer Three

• The study did not show any differential treatment effect in primary outcome, improvement in HbA1c, though both groups showed very impressive improvement both at 6- and 12-month. In addition, none of the secondary clinical outcomes showed any differential treatment effect except for SBP, but due to very small magnitude of difference and looking at a large number of secondary outcomes, this finding might be attributed to large type I error.

We agree that the significant differential treatment effects observed for the secondary outcomes may be attributed to type I error. In pragmatic trials it is very difficult to control all sources of variance (see response to Reviewer 1 bullet points 10 and 11). This means that it can be difficult to see signals that the intervention is working. This has been added into the limitations sections in the discussion (page 31, para 2):

"Another limitation is that only small differential effects in some secondary measures were detected, and although small effects in pragmatic trials are common since all sources of variance cannot be controlled, it is possible that they result from type 1 errors."

• The authors claimed that they observed a significant improvement in intervention subjects compared to control subjects in psychological distress outcome, but none of the two groups showed any improvement in psychological distress, table 2 shows no significant change at 6-month and a small magnitude of worsening score for the intervention group and a large magnitude of worsening score for the control group at 12-month.

Thank you for pointing this important detail out. The text in the abstract and main text has been amended accordingly.

Abstract (results section): "Psychological distress also significantly worsened in the control group compared to the intervention group (adjusted difference -2·66, 95% CI -4·97 to -0·35, p=0.02)."

Main Text: When discussing psychological distress throughout the main text we now refer to the significant difference in statements such as: "psychological distress significantly worsened in the control group"; "greater psychological distress in the control group compared to the intervention". Any reference that suggests the intervention group displayed improvement in psychological distress has been amended.

• Though the authors claimed there might be a dilution treatment effect due to possible contamination or drawing similar attention to the control subjects, it should not be concluded that intervention was better than the usual care.

The conclusive statement in the abstract (bottom page 3) and main text (conclusion - page 32, para 2) has been amended to remove any reference that concludes that the intervention was better than

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

usual care.

Abstract:
“In line with the current Chinese primary health care reform, this study is the first large-scale cluster RCT to be implemented within real world CHSs in China, specifically addressing T2DM. Although a differential treatment effect was not observed for HbA1c, some secondary T2DM-related outcomes improved significantly in the intervention group. Additionally, numerous outcomes (including HbA1c) improved in both groups, supporting the establishment of regular, free clinical health checks for people with T2DM in China.”

Main Text:
“The coaching approach employed in this study is entirely consistent with China’s primary health care reforms. T2DM is one of China’s most prevalent and burdensome chronic illnesses and one that is amenable to effective community based primary health care interventions. Significant changes within both groups with regard to numerous clinical outcomes, including HbA1c, highlight the advantages of regular, free clinical health checks for patients with T2DM. With a strong focus on self-management and health coach support, this study has the potential to be adapted to other chronic diseases, as well as other regions of China. Indeed, an extension of the present trial will be implemented in Shenzhen in 2016/17 as part of the functions of the newly formed International Primary Health Care Research Institute.”

• While there was very impressive improvement in HbA1c for both groups, there was either no or very minimal improvement in fasting glucose. Can the authors provide any explanation for such observation?

Thank you for raising this point. We note that FPG and HbA1c typically correlate very reliably with each other in the management of T2DM, however due to the complex relationship between the two we are unable to accurately explain why the change in HbA1c was not reflected in FPG in our study. We have added this as a discussion point though (page 27, para 2).

VERSION 2 – REVIEW

REVIEWER	Peter Rohloff Wuqu' Kawoq | Maya Health Alliance - Guatemala Brigham and Women's Hospital - USA
REVIEW RETURNED	27-Jan-2016

GENERAL COMMENTS	I think that the paper is much stronger now, with a lot of important details added that were requested by the three reviewers. In addition, both the strengths and weaknesses of the research are better highlighted. The paper is much more pragmatic now, which I think was the main missing element before, and therefore will be of great utility for other DM implementers around the world. At this point, I have only minor queries, some of which were introduced by the edits: - Page 13 Line 12: I fear I’m now unclear on what “payment” usually entails and how this differed in this trial. Which were the waived fees? Are we talking about medical consultation fees, or did patients in this trial receive free medications or lab tests that otherwise were unavailable in the region? Did the trial pay for these, or where did this funding come from? - inclusion of Sensitivity Analysis and discussion of missing data is

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

	helpful and makes the paper much stronger, thanks for adding all of these details - Table 1: “The intervention and control groups were similar for all variables.” Can this statement to be upgraded to “no statistical difference between groups” on these variables? Potentially also with a notation about statistical significance in the legend of Table 1 also, for those readers who skip text and go right to the tables and figures. - Table 2 is much more readable, thanks - Table 3 is very valuable, thanks for including - Discussion of psychological distress clinical significance is great, thanks for adding -Page 27: discussion of discrepancy between A1C and fasting glucose: Another possible explanation, are we sure from monitoring protocol adherence that these fasting glucose results are indeed fasting? -Page 28, the discussion of altering adherence by incentivizing through free care comes up again, and the discussion is strong. Not knowing the China context, it does make me wonder again what people do/do not pay for under true “usual care” context (meds, labs, visits). A few more words on this here or in the introductory sections would be helpful. - Something to consider for the results or discussion section: any ideas about why the QOL assessments worsened in the trial?

REVIEWER	Lise Juul Aarhus University, Department of Public Health, Denmark
REVIEW RETURNED	22-Jan-2016

GENERAL COMMENTS	The paper has improved a lot, and I only have a few comments; It is a very good example of a very complex intervention. Actually, the training of the health coaches could have been described as a part of the intervention on an organizational level according to the socio-ecological model. I think it is a bit incorrect to categorize SDSCA as psychosocial outcomes. The risk of multiple significance could have been discussed. A discussion on the representativity of the study population is lacking. The magnitude of the reductions in HbA1c needs more discussion. Have other studies found similar improvements? Finally, the conclusion does not exactly answer the objective of the study, which should be added.

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

REVIEWER	Chandan K. Saha Indiana University School of Medicine
REVIEW RETURNED	08-Jan-2016

GENERAL COMMENTS	Thanks for addressing the concerns I raised and making appropriate changes.

VERSION 2 – AUTHOR RESPONSE

• It is a very good example of a very complex intervention. Actually, the training of the health coaches could have been described as a part of the intervention on an organizational level according to the socio-ecological model.

Thank you for your comment. We note that the training of health professionals could be described within the organisational level of the socio-ecological model. We have opted not to introduce the discussion of this model and its multiple levels of influence into the main text as we feel it would alter the intended scope of the current paper.

• I think it is a bit incorrect to categorize SDSCA as psychosocial outcomes.

SDSCA has now been referred to throughout the manuscript as a measure of self-care behaviour.

• The risk of multiple significance could have been discussed.

We agree and have now addressed this issue within the discussion, by noting the possibility of Type I errors (bottom page 31):

"...Another limitation is that only small differential effects in some secondary measures were detected, and although small effects in pragmatic trials are common since all sources of variance cannot be controlled, it is possible that they result from type I errors. Multiple examinations also increase risk of type I errors."

• A discussion on the representativity of the study population is lacking.

Although we randomly selected patients, the lack of accurate recruitment records for those who declined to participate consequently limited our ability to comment on the representativity of our sample. However, we have compared the baseline clinical and demographic characteristics of our sample with a recently published prevalence study and have inserted the following statement within the discussion section of the manuscript (page 30):

"Indeed, the baseline demographic and clinical characteristics of our sample were similar to that previously observed in individuals with DM who live in urban and developed regions in China. For example, in our sample there were approximately similar proportions of males and females; the mean age of our sample (64 years) was within the second highest population prevalence age band for DM (slightly higher population prevalence for ≥70 years); participants were more likely to have education levels of secondary/high school or higher, and were also more likely to be overweight, and have elevated total cholesterol, LDL cholesterol, and triglyceride levels."

• The magnitude of the reductions in HbA1c needs more discussion. Have other studies found similar improvements?

No other studies have similarly observed the magnitude of reductions in HbA1c. The following

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

sentence has been added to the discussion section (page 30):

“However, no previous studies have observed improvements in HbA1c of the same magnitude as the present trial. The baseline HbA1c values of these studies were substantially lower (~7%), consequently limiting their potential for improvement.”

• Finally, the conclusion does not exactly answer the objective of the study, which should be added.

The concluding paragraph has been updated to include a statement on the effectiveness of the present study at improving the primary and secondary outcomes, compared to usual care (page 32):

“No differential treatment effects were observed for HbA1c or the majority of secondary outcomes, however significant changes within both groups with regard to numerous clinical outcomes, including HbA1c, highlight the advantages of regular, free clinical health checks for patients with T2DM.”

• Page 13 Line 12: I fear I’m now unclear on what “payment” usually entails and how this differed in this trial. Which were the waived fees? Are we talking about medical consultation fees, or did patients in this trial receive free medications or lab tests that otherwise were unavailable in the region? Did the trial pay for these, or where did this funding come from?

To make the issue of payment clearer, the paragraph in the methods section now reads (page 13, para 3):

“Participants in both treatment groups did not receive payment for participation in this study. However, medical fees (both consultation and out-of-pocket pathology fees) associated with participation in the project were waived, with the associated costs absorbed by CHSs.”

• Table 1: “The intervention and control groups were similar for all variables.” Can this statement to be upgraded to “no statistical difference between groups” on these variables? Potentially also with a notation about statistical significance in the legend of Table 1 also, for those readers who skip text and go right to the tables and figures.

Thank you for this suggestion. We can assure the reviewer that there were no significant differences found between the baseline groups - see p values for each baseline characteristic below. However, we have not included these p-values or made reference to baseline statistical tests within the manuscript because the CONSORT statement recommends against doing this since any differences would have resulted from chance only as randomization was applied (see http://www.consort-statement.org/checklists/view/32-consort/510-baseline-data). However, we have mentioned the similarity of baseline factors between the two groups in the text in the results section (page 15):

“The intervention and control groups were similar for all variables at baseline, and even if statistical differences between groups were observed then the analysis method could have accounted for this by adjusting for baseline scores.”

Baseline characteristics
Age in years, 0.539
Female, 0.479
Married (including de facto), 0.508
Retired, 0.736
Secondary/high school education, 0.238
Duration of T2DM in years, 0.427
Currently prescribed insulin, 0.282

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.

Co-morbid conditions present, 0.051
Current Smoker, 0.636

• Page 27: discussion of discrepancy between A1C and fasting glucose: Another possible explanation, are we sure from monitoring protocol adherence that these fasting glucose results are indeed fasting?

Yes, we can be sure that the fasting glucose results were indeed fasting. Fasting time was asked of participants and recorded on health check documentation prior to each health check. If participants had not fasted, they were asked to reschedule their appointment and were reminded of the appropriate fasting duration. The following statement has been added to the methods section to clarify this point (page 10):

“All participants were instructed to fast overnight for a minimum of eight hours, and participant fasting times were recorded prior to each blood test. Where fasting times were not sufficient, participants were asked to reschedule their appointment.”

Additionally, we have added an additional sentence in page 27 for the discussion of the discrepancy between HbA1c and FPG:

“We can exclude the possibility of non-fasting blood being sampled as the fasting duration was documented at each health check.”

• Page 28, the discussion of altering adherence by incentivizing through free care comes up again, and the discussion is strong. Not knowing the China context, it does make me wonder again what people do/do not pay for under true “usual care” context (meds, labs, visits). A few more words on this here or in the introductory sections would be helpful.

As per bullet point 7, the amended paragraph regarding payment in the methods section is now clearer to include the typical out of pocket expenses that are associated with usual care (page 13):

“Participants in both treatment groups did not receive payment for participation in this study, however, medical fees (both consultation and out-of-pocket pathology fees) associated with participation in the project were waived, with associated costs absorbed by CHSs. Although China has near universal health insurance coverage, individuals with T2DM typically incur out-of-pocket expenses for both medical (consultation and pathology fees) and pharmaceutical care.”

• Something to consider for the results or discussion section: any ideas about why the QOL assessments worsened in the trial?

The results section notes the variability of the within-group significant findings observed between groups for the different QoL domains. Due to the large variability between domains and treatment groups, we have opted not to further discuss the QoL assessment within the discussion.

BMJ Open: first published as 10.1136/bmjopen-2015-009319 on 4 March 2016. Downloaded from http://bmjopen.bmj.com/ on April 19, 2024 by guest. Protected by copyright.