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)	A novel education-based intervention to reduce inappropriate antibiotic prescribing for treatment of gonorrhea in China: protocol for a cluster randomized controlled trial
AUTHORS	Jiang, Ting-Ting; Yang, Yun-Qing; Cao, Ning-Xiao; Yin, Yue Ping; Chen, Xiang-Sheng

VERSION 1 - REVIEW

REVIEWER	Joe Hicks University of Leeds, U.K.
REVIEW RETURNED	13-Feb-2020

GENERAL COMMENTS	Dear authors, Although the core information required for a protocol is mostly present, currently there are a lot of important details missing that are required for a fully informative and robust protocol. See my comments for suggestions about how to make the protocol more informative and more robust. Please be aware that not addressing these issues could well lead to strong criticism at peer review. In particular, many of the details currently missing mean that you are leaving open the possibility that you could innocently or intentionally manipulate the results of the study through your analysis, and this is exactly what a protocol should aim to prevent. As explained below you may wish to produce a separate statistical analysis plan to fully detail your analysis plans. All the best with the study. It sounds very interesting and important. ### Abstract ### Define what your clusters are. State how many provinces. Specify the total no. clusters per arm. Consider specifying the stratification characteristics (less critical). Specify that the primary outcome is at the cluster level. Secondary outcome: see comments on outcomes section below and update either with a brief statement about the types of secondary outcomes you will measure, or you could just remove this entirely as your effectiveness will be judged on the primary outcome. ### Introduction ### Very well written and clear outline of the problem and aim of the study. However, it would benefit from very briefly describing what the intervention involves as part of the description of the overall aim of

	the study, and outlining what studies, or if any studies, have looked at the intervention previously (not necessarily in relation to appropriate AB prescribing for this disease), and what the current evidence appears to show (i.e. a brief review of the existing evidence). ### Methods ### ### Study design: you must fully and clearly define what your clusters are, I would recommend doing this in the study design section for ease of understanding for readers rather than only revealing this later on. Also, add more detail on the types of facilities that are the clusters, e.g. what hospital level are they, are they only rural, urban or both, e.g. township hospitals, community health centres? Clarify how the internal pilot area is selected (e.g. purposively [and if so on what criteria] or randomly [if randomly how?]), and how many clusters will be in the internal pilot, and make it clearer which trial arm. It appears to only be in the intervention arm, but be aware that this then means that those clusters will be "out of synch" with the remaining intervention and control clusters and this may lead to bias in the results. Ideally any internal pilot intervention clusters should have randomly allocated control clusters from which you collect data at the same time, so across the study as a whole all "groups" of intervention and control clusters are matched in terms of the data collection periods, as there may be underlying temporal changes/effects that will bias the results otherwise. ### Study sites: info here can either be merged with study design section or will need expanding with cluster definition info as explained above. More info could be added on the types of areas the facilities are located in. e.g. Guangxi and Shanghi provinces differ substantially. Also, some info on how provinces were selected would be useful. ### Sample size 1) this section should probably come below the outcomes section, because you need to understand the outcomes before understanding the sample size; 2) clarify is the 10% difference the expected increase or a minimum improvement that you want to be able to detect (the recommended approach is to choose a minimum and not base on an expected change, as this is often overly optimistic), and it should be clear what criteria were used to decide this figure, or what processes (e.g. consultation with relevant experts and stakeholders?); 3) explain what k is based on, e.g. literature/other studies/pilot data/guess?; 4) explain what n is based on, e.g. minimum number of expected patients per cluster?, because if this is likely to vary and sometimes be below 100 (appears possible given data collection info) then you need to explicitly account for this variability in cluster size in the sample size calculation or it will be biased too low (a conservative option is to just choose the minimum possible cluster size); 5) it needs to be clear how the hospitals will be selected within each province (e.g. simple random sample from the total list across all provinces, stratified random sample by province and/or other factors, purposive sample [and if so on what criteria] etc). ### Cluster randomisation 1) fully define the characteristics and cut-points used to create the

	stratification groups for hospital randomisation, and how the internal pilot randomisation works within this randomisation process; 2) clarify is the stratified randomisation simple randomisation within each strata?; 3) masking - what about patients, will/can they be masked, and very importantly what about data collectors/processors (later on you mention prescriptions are screened for inclusion – there is a big risk of bias here if not masked)? ### Eligibility for participation: what are the patient eligibility criteria? There must be at least one, i.e. have gonorrhea, and if there are any others state them clearly here. ### Intervention and control: 1) clarify/explain more what the interaction between a public health staff and physician involves and aims to achieve in terms of "interpreting" the intervention, as it's not really clear what this means in practical terms; 2) strongly consider using the TIDIER guidelines (https://www.equator-network.org/reporting-guidelines/tidier/) to guide your explanation of the intervention details as there is important information lacking, e.g. who provides Drs with the info card, who explains the intervention to them (aims etc), do all participating Drs get the card and info at the same time?; 3) you must be clear how long will the intervention run for? ### Outcomes Primary outcome is not fully defined. 1) What prescriptions are included in the outcome (presumably those given to eligible patients with gonorrhea during the survey periods, or will just a subset of the total number of prescriptions issues [i.e. 100 as per the sample size] be collected per cluster during each survey, and if so how will these be selected from the total number - randomly, just the first 100 etc?), 2) it needs to be clear what level the outcome is derived at, e.g. I assume it's the cluster-level or the per cluster proportion of prescriptions?, 3) it needs to be clearer when these outcomes will be collected. Baseline and evaluation surveys are mentioned but it should be more explicit and clear for each survey how long the data will be collected for. I see you detail the data collection periods in the data collection section, but it makes it easier to understand if you explain the data collection periods in the outcomes section and either repeat this in the data collection section if you feel the need or just talk about baseline and evaluation surveys without explicitly mentioning the periods again. The "secondary outcome" is not defined at all. From reading further on it appears that you will be collecting a wide range of data from physicians via a questionnaire on likely reasons for non-adherence, but this means there will be lots of possible secondary outcomes you could define and analyse. However, you cannot refer to all such data as “the secondary outcome”. You must explicitly and fully define every single outcome that you intend to analyse from the trial in terms of the effect of the intervention on it, just like for the primary outcome. If you don’t define your outcomes fully and clearly in the protocol you cannot claim they were pre-planned in your results paper, and they will rightfully be viewed sceptically/criticised as not pre-planned and open to investigator manipulation in terms of choosing outcomes that show desired results. You need to think about all the data you plan to collect and just like for the primary outcome you should state what type of variable the secondary outcome or outcomes are (e.g. binary, proportion, continuous,

	mean), what level they are derived at (individual/cluster), what units or categories they have (e.g. unaware of national guidelines = yes/no), and state how they will be measured and when (e.g. questionnaire at baseline and endline or just endline?). What about harm related outcomes? Are these not necessary? You should justify why if not. ### Data collection 1) When is the physician questionnaire data collected? Baseline and evaluation or just basline? 2) The data collection periods are not clear in relation to the baseline and evaluation surveys mentioned. It seems to imply there is just one data collection period from one month after implementation until at least 6 months after, so where does the baseline survey occur? Is the baseline survey just the "baseline data of prescribing adherence will be retrospectively collected from the most recent 30 archived prescriptions prior to the trial in each of the departments in each of participating hospitals using"? I think it could be explained more explicitly, e.g. "we will collect all outcomes and covariates at baseline and endline, with the baseline survey being... and the endline survey being ..." Also it could be clearer if the primary and/or second outcome data is collected at baseline or is this just the secondary data? 3) You say the data will be "conducted monthly thereafter for at least 6 months". This is not advisable because it allows subjective decision making about how long data are collected for, and therefore opens the door for researchers to influence data collection processes and trial results. This could be criticised later on as questioning the robustness of the processes and results. The data collection period should be fixed and based on good information about the time needed to collect the necessary number of prescriptions. 3) Where does the prescription data come from? Electronic routine records, or paper prescriptions? 4) Who screens the prescriptions to check they are eligible? This must be clear and how independent decisions are maintained, especially around masking? i.e. will screening individuals be blinded or not? 5) What about covariate data, e.g. will you collect data from prescriptions on age, sex etc? List all variables that will be collected. Do the same for any covariates collect at the physician level and/or hospital level, e.g. Dr age, sex, experience/level, hospital size? Or refer to where these data are collected if from the tools in the appendix. ### Data analysis There is a lot of detail missing currently that will open your results up to strong criticism. This can be added here or included in a separate statistical analysis plan. See comments below for details. 1) You appear to suggest you will test for baseline differences in characteristics with null hypothesis significance testing (chi-sq, t-test and Mann Whitney U specifically). This is not recommended and explicitly discouraged by CONSORT and more generally as the null hypothesis is known to be true (because the groups only differ due to random sampling as they were randomly assigned), so the p-value is completely meaningless. You only need to present descriptive statistics of characteristics for each group, i.e. frequencies and percentages, means and SDs or medians and IQRs etc. 2) You must state clearly the priority of your analyses in terms of which will be used as the primary analysis to determine

	effectiveness, i.e. the crude or covariate adjusted analysis? As with the secondary outcomes if you do not specify which of these analyses will be used to make the determination about treatment effectiveness then you are leaving open the possibility that you can choose between crude or adjusted analyses as to which gives you the most favourable result. I would suggest your adjusted results are used with the crude results acting simply as a sensitivity analysis to illustrate whether covariate adjustment is important for the results or not. Adjusting increases power and reduces the effect of imbalances between arms with typically minimal risks. 3) You need to be clear about which outcomes you are describing the analysis of (explicitly state), and what effect measures you will estimate, e.g. if the primary outcome then as a proportion it will be a risk difference or risk ratio depending on which approach you take, but this should be clear and isn't currently. The p-value is not the key measure of effect, that will be the size of effect and 95% CI around that estimate (e.g. risk difference). Similarly, for your secondary outcomes all analyses should be described in the same level of full detail. You also must fully specify which covariates will be adjusted for in your adjusted analyses, and what units and categories those covariates have (e.g. age [years], education [none, primary, secondary, further] etc). This is particularly important in terms of being clear whether the baseline outcome data will be used as a covariate. Again not doing this opens your adjusted analyses up to criticism that you chose covariates that gave you desirable results after collecting the data. 4) You have not explained how you will deal with missing data. This needs a full treatment. E.g. will you use any form of imputation (if so provide full details), or just complete case analysis, and if so what are the justifications for doing so? 5) You do not specify any subgroup analyses. If you do not plan to do any this should be explicitly stated. If you do plan to do any they should be fully detailed (all methods) here. Again if you do not do this then you cannot call them pre-planned in your results paper and they will rightly be viewed as very open to bias (even more than subgroup analyses already are) because again it leaves open the possibility that you selected analyses that suited what results you wanted to find. 6) CONSORT recommends avoiding using the term "intention to treat" because it has no universally accepted definition. Instead you should explicitly describe who will be in your main and any sensitivity analyses. It is particularly important to consider how any missing data and how you deal with that will affect these analysis populations. E.g. if you only do complete case analyses and have some missing data then your analysis populations will not be the full set of originally "recruited" patients whose data you collected. Therefore, it should be completely clear who will be included in the analyses, and any sensitivity analyses (e.g. "per protocol" analyses, which again is a fairly meaningless term and should instead be fully defined in terms of who what will include). Refer to your trial flow chart to help readers understand this better. However, the flow chart needs to be clear whether the internal pilot sites are only intervention sites and if the data collected in the control arm started after the pilot. Consider the relative placement of the boxes as the control arm one is in between the intervention pilot and main trial box which is ambiguous. 7) You say single and multiple regression analyses will be further conducted to identify factors related to nonadherence to the National Guidelines among participants in the intervention arm. This may be better for a subsequent paper than your main trial results paper, but that is of course up to you, however these methods are not well defined again as they are just a few words. First, note that only multi-level regression methods would be suitable

	given the clustered nature of the data, second, what covariates will be used, third, how will you deal with missing data etc (you need to provide the same details for all analyses). Lastly, it's unclear why you would do single and multiple regression analyses, given that single regression analyses will just give you more confounded results than multiple regression analyses? 8) State if there are any interim analyses planned or not, and if so provide full details. ### Ethics and dissemination: I'd move the consent process details into the methods as it's easier to understand there when explained as part of the recruitment and randomisation process, also it needs to be clear who consents for hospitals to be part of the study, and it should be clear if patients are not asked for consent and if so this should be justified. It should also be clear who recruits hospitals and physicians into the study, e.g. research team members? Again I would say all this detail should go into the methods under an appropriate section (e.g. "recruitment and randomisation" would make sense because then you can explain how these things happen chronologically).

REVIEWER	Sun Qiang School of Healthcare Management, Shandong University, P.R. China
REVIEW RETURNED	19-Feb-2020

GENERAL COMMENTS	The manuscript plans to use a cluster randomized controlled trial to evaluate the effects of an education-based intervention on improving the rational use of antibiotics. the design of cluster randomized controlled trail is clearly described and the process evaluation was also considered, the outcome measurements are acceptable. But the key problems of the manuscript is the intervention itself. How can the authors justify the novelty of the education-based intervention? What are the biggest advantages of this video learning materials compared to the traditional printed training booklets? How do the study ensure the participates to look at the video as required? Please justify them. In addition, why the study randomly select 17 prescriptions per months in each hospital? how was the sample size calculated.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Joe Hicks

Institution and Country: University of Leeds, U.K.

Although the core information required for a protocol is mostly present, currently there are a lot of important details missing that are required for a fully informative and robust protocol. See my comments for suggestions about how to make the protocol more informative and more robust. Please be aware that not addressing these issues could well lead to strong criticism at peer review. In particular, many of the details currently missing mean that you are leaving open the possibility that

you could innocently or intentionally manipulate the results of the study through your analysis, and this is exactly what a protocol should aim to prevent. As explained below you may wish to produce a separate statistical analysis plan to fully detail your analysis plans. All the best with the study. It sounds very interesting and important.

Response:

Thanks for these comments and suggestions. We try out best to address all the concerns and make updates accordingly in the revised manuscript.

Abstract

Define what your clusters are. State how many provinces. Specify the total no. clusters per arm. Consider specifying the stratification characteristics (less critical). Specify that the primary outcome is at the cluster level. Secondary outcome: see comments on outcomes section below and update either with a brief statement about the types of secondary outcomes you will measure, or you could just remove this entirely as your effectiveness will be judged on the primary outcome.

Response:

We had almost all these information in the previously submitted version but we are happy to reword the statement to read as “We will conduct a cluster randomized control trial (cluster-RCT) across 144 hospitals (clusters) in 8 provinces in China.” (Page 2)

We have reorganized the statements of the primary and secondary outcomes in the revised manuscript to read as “The primary outcome is to estimate the proportion of prescriptions adherent to the regimens (1g intramuscular ceftriaxone single dose) recommended by the National Guidelines for treatment of uncomplicated gonorrhea at the cluster level. In addition, to understand physician’s barriers to the intervention by conducting a questionnaire survey will be considered as the secondary outcome of the study.” (Page 2)

Introduction

Very well written and clear outline of the problem and aim of the study. However, it would benefit from very briefly describing what the intervention involves as part of the description of the overall aim of the study, and outlining what studies, or if any studies, have looked at the intervention previously (not necessarily in relation to appropriate AB prescribing for this disease), and what the current evidence appears to show (i.e. a brief review of the existing evidence).

Response:

The intervention through an on-line video-based training programme is a critical part of the study to validate the effectiveness and feasibility of such intervention to improve the current adherence of clinical practitioners to national treatment guidelines in China. In the revised manuscript, we have added the following description to briefly address this: “we will conduct a prospective cluster randomized trial with a primary objective to observe whether the proposed intervention (on-line video-based training programme) could significantly improve physicians’ prescribing behaviors to adhere to the up-to-dated National Guidelines25 in treatment of uncomplicated gonorrhea (1g intramuscular ceftriaxone single dose).” (Page 6)

We have outlined the previous studies in which mobile technology-based training programmes were applied as an intervention and summarized the currently available evidence as the following

statement: “Mobile health (mHealth), defined as interventions and programmes designed to support medical and public health through the use of mobile technology, has been used as a tool to deliver continuous training and education to healthcare providers.20-21 Previous randomized controlled trials (RCTs) in some medical fields have revealed the success of mHealth in optimizing antibiotic prescribing.22-24” (Page 6)

Methods

Study design: you must fully and clearly define what your clusters are, I would recommend doing this in the study design section for ease of understanding for readers rather than only revealing this later on. Also, add more detail on the types of facilities that are the clusters, e.g. what hospital level are they, are they only rural, urban or both, e.g. township hospitals, community health centres? Clarify how the internal pilot area is selected (e.g. purposively [and if so on what criteria] or randomly [if randomly how?]), and how many clusters will be in the internal pilot, and make it clearer which trial arm. It appears to only be in the intervention arm, but be aware that this then means that those clusters will be “out of synch” with the remaining intervention and control clusters and this may lead to bias in the results. Ideally any internal pilot intervention clusters should have randomly allocated control clusters from which you collect data at the same time, so across the study as a whole all “groups” of intervention and control clusters are matched in terms of the data collection periods, as there may be underlying temporal changes/effects that will bias the results otherwise.

Response:

Our study considers the participating hospitals as clusters for this cluster RCT and the corresponding statement in the revised manuscript reads as “Hospitals will be used for randomization of them into intervention or control arm,” (Page 7)

We purposively (rather than randomly) recruit the hospitals to participate in the study because we intend to capture those hospitals with high volume of reported cases of gonorrhea. Based on our national case-reporting system, these hospitals are mostly located in urban cities. We reword the statement to read as “Within each study province, we will invite the city-level hospitals with the high volume of reported gonorrhea cases to participate in the study. The hospitals at national or province-level in the study provinces will be excluded because small number of such hospitals within a province makes a difficulty to ensure comparable assignments of these hospitals into intervention and control arms.” (Page 7)

We planned to purposively select one of the 8 provinces for the internal pilot trial but we are happy to take the reviewer’s comment to use a randomized process to select one province for the pilot trial. The selection process has been updated in the revised manuscript to read as “The internal pilot trial will be conducted in one province randomly selected from the 8 study provinces to collect the data simultaneously in 9 intervention and 9 control hospitals during a period of one month for examining feasibility and acceptability of the intervention.” (Page 7)

Study sites: info here can either be merged with study design section or will need expanding with cluster definition info as explained above. More info could be added on the types of areas the facilities are located in. e.g. Guangxi and Shanghi provinces differ substantially. Also, some info on how provinces were selected would be useful.

Response:

Thanks for the reviewer's valuable suggestion. We have merged the study sites section with study design section as suggested. (Page 6-7)

Selection of provinces for the study is based on the reported incidence of gonorrhoea in 2018. The 8 provinces (Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong, Guangxi, Hainan and Yunnan) are purposively selected from 31 provinces in mainland China in consideration of their higher incidence of reported gonorrhoea cases than the national average (Figure 2). It is true that substantial heterogeneities exit across provinces (e.g. Guangxi and Shanghai). However, such heterogeneities may favor representativeness of the study findings to the whole country but not compromise the comparability between intervention and control arms because the study clusters to be assigned into intervention and control arm are hospitals within a province rather than the provinces.

Sample size

1) this section should probably come below the outcomes section, because you need to understand the outcomes before understanding the sample size;

2) clarify is the 10% difference the expected increase or a minimum improvement that you want to be able to detect (the recommended approach is to choose a minimum and not base on an expected change, as this is often overly optimistic), and it should be clear what criteria were used to decide this figure, or what processes (e.g. consultation with relevant experts and stakeholders?);

3) explain what k is based on, e.g. literature/other studies/pilot data/guess?;

4) explain what n is based on, e.g. minimum number of expected patients per cluster?, because if this is likely to vary and sometimes be below 100 (appears possible given data collection info) then you need to explicitly account for this variability in cluster size in the sample size calculation or it will be biased too low (a conservative option is to just choose the minimum possible cluster size);

5) it needs to be clear how the hospitals will be selected within each province (e.g. simple random sample from the total list across all provinces, stratified random sample by province and/or other factors, purposive sample [and if so on what criteria] etc).

Response:

1) We are happy to take the suggestion to move sample size section below the outcome section. (Page 9-10)

2) The 10% difference as a minimum adherence improvement expected from the intervention will be conservatively deemed for clinical relevance based on a consultation of STD clinical experts and control programmers. The corresponding statement has been updated in the revised manuscript to read as "based on a consultation of STD clinical experts and control programmers, a minimum 10% increase in adherence to the National Guidelines is deemed clinically relevant in the intervention arm." (Page 9)

3) & 4) The value of k is based on a nationwide survey in China and n = 100 represents the minimum number of uncomplicated gonorrhea patients per hospital within six months. The statements read as "n = 100 (minimum number of uncomplicated gonorrhea patients per hospital within six months), and k = 0.1 (based on a nationwide survey18 in China)." (Page 9)

5) We have responded the similar comment above to explain the processes to select study provinces and hospitals within each study province.

Cluster randomization

1) fully define the characteristics and cut-points used to create the stratification groups for hospital randomisation, and how the internal pilot randomisation works within this randomisation process;

2) clarify is the stratified randomisation simple randomisation within each strata?;

3) masking - what about patients, will/can they be masked, and very importantly what about data collectors/processors (later on you mention prescriptions are screened for inclusion – there is a big risk of bias here if not masked)?

Response:

1) Sorry for not including the detailed description in our previous manuscript. Actually, we will stratify the hospitals (clusters) into two groups (provincial capital city and other cities) within each study province, and 4 hospitals from capital city and 14 hospital from other cities were purposively selected in consideration of a high volume of reported cases. In each stratification, a simple randomisation process will be used to assign the hospitals into intervention or control arm. Regarding the province to select for pilot internal trial, a simple randomisation process will be used. We have updated the description to read as “Within each province, we will stratify the hospitals (clusters) into two groups (provincial capital city and other cities), and 4 hospitals from capital city and 14 hospital from other cities will be purposively selected in consideration of a high volume of reported cases. In each stratification, a simple randomization process will be used to assign the hospitals into intervention or control arm. Thus, we will have 72 hospitals assigned to the control group and 72 to the intervention group, 9 and 9 respectively for each one of the 8 provinces participating in our study. After all 144 hospitals were randomized, one province will be randomly selected from the 8 participating provinces to become the internal pilot area using closed and opaque envelopes by an independent person from the research team, thus the 18 participating hospitals (9 from each arm) in this selected province will become the internal pilot clusters. The remaining 126 hospitals (63 from each arm) will, therefore, participate in the main trial, along with the 18 hospitals involved in the internal pilot (Figure 1).” (Page 10)

2) This has been clarified in the point 1).

3) Masking is a critical issue for RCT but the trial will be not possible to blind either in-hospital project coordinators or participating physicians on the arm assignment, given the explicit nature of the intervention components (e.g. physicians in the intervention hospitals will be trained by in-hospital project coordinators). However, the prescription assessors and data analyst will not be aware of the assignments. We added this statement into the revised manuscript. (Page 10-11)

Eligibility for participation: what are the patient eligibility criteria? There must be at least one, i.e. have gonorrhea, and if there are any others state them clearly here.

Response:

There are two eligibilities for participation (physician and prescription). So we added a paragraph in the revised manuscript to read as “Physicians who work in the out-patient departments of dermatology (dermatovenerology), urology, andrology or gynaecology and provide informed consent (see Supplement 1) will be eligible for participating the study. Prescriptions of these physicians for treatment of uncomplicated gonorrhea will be eligible for evaluating the adherence to the National Guidelines,” In addition, as suggested by the Editorial office above we added the exclusion criteria. (Page 8)

Intervention and control

1) clarify/explain more what the interaction between a public health staff and physician involves and aims to achieve in terms of "interpreting" the intervention, as it's not really clear what this means in practical terms;

2) strongly consider using the TIDIER guidelines (https://www.equator-network.org/reporting-guidelines/tidier/) to guide your explanation of the intervention details as there is important information lacking, e.g. who provides Drs with the info card, who explains the intervention to them (aims etc), do all participating Drs get the card and info at the same time?;

3) you must be clear how long will the intervention run for?

Response:

1) In China, each hospital has at least one public health staff to be responsible for case-reporting of STIs (mainly syphilis and gonorrhea) and other infectious diseases and also responsible for training of medical staff on STIs. Traditionally, these public health staff are usually requested to participate in the training-of-trainers workshop organized by local health authorities or disease control department and then organize a one-off training hopefully participated by all relevant physicians in their own hospitals but such training does not work very well. This is actually a supposed practice in the control arm in our study. We added a description to read as “In the intervention arm hospitals, project coordinators will provide a less than 30-minutes interactive training session in each of the relevant departments (integrated within monthly department meeting) prior to implementation of the intervention to interpret the information card and the way to use the QR code for accessing to the on-line training video.” (Page 9)

2) We have explained the intervention details according to the TIDIER guidelines (see supplementary file).

3) The intervention will last for 6 months during which the participating physicians will be regularly reminded by the project coordinator. We have added a paragraph to read as “The intervention of video-based training will last for 6 months and during the 6 months the participating physicians will be reminded for video playback by the project coordinators through sending a WeChat message monthly.” (Page 9)

Outcomes

Primary outcome is not fully defined.

1) What prescriptions are included in the outcome (presumably those given to eligible patients with gonorrhea during the survey periods, or will just a subset of the total number of prescriptions issues [i.e. 100 as per the sample size] be collected per cluster during each survey, and if so how will these be selected from the total number - randomly, just the first 100 etc?),

2) it needs to be clear what level the outcome is derived at, e.g. I assume it's the cluster-level or the per cluster proportion of prescriptions?,

3) it needs to be clearer when these outcomes will be collected. Baseline and evaluation surveys are mentioned but it should be more explicit and clear for each survey how long the data will be collected for. I see you detail the data collection periods in the data collection section, but it makes it easier to understand if you explain the data collection periods in the outcomes section and either repeat this in the data collection section if you feel the need or just talk about baseline and evaluation surveys without explicitly mentioning the periods again.

Response:

Similar to our response to the above comment on primary outcome, we have further defined it to read as “The primary outcome is the proportion of prescriptions adherent to the regimens for treatment of uncomplicated gonorrhea recommended by the National Guidelines at the cluster level.” (Page 9)

1) Regarding the prescriptions to include for measuring the outcome, “The most recent 100 eligible prescriptions will be obtained by the prescription assessors from each study hospital prior to implementation of the intervention as baseline data and 6 months after the implementation for evaluating the outcome.” (Page 9)

2) The primary outcome will be evaluated at the cluster level It is has been updated accordingly on Page 9.

3) The data will be collected prior to implementation of the intervention and then 6 months after the implementation. This comment is relevant to our response to point 1).

The "secondary outcome" is not defined at all. From reading further on it appears that you will be collecting a wide range of data from physicians via a questionnaire on likely reasons for non-adherence, but this means there will be lots of possible secondary outcomes you could define and analyse. However, you cannot refer to all such data as “the secondary outcome”. You must explicitly and fully define every single outcome that you intend to analyse from the trial in terms of the effect of the intervention on it, just like for the primary outcome. If you don’t define your outcomes fully and clearly in the protocol you cannot claim they were pre-planned in your results paper, and they will rightfully be viewed sceptically/criticised as not pre-planned and open to investigator manipulation in terms of choosing outcomes that show desired results. You need to think about all the data you plan to collect and just like for the primary outcome you should state what type of variable the secondary outcome or outcomes are (e.g. binary, proportion, continuous, mean), what level they are derived at (individual/cluster), what units or categories they have (e.g. unaware of national guidelines = yes/no), and state how they will be measured and when (e.g. questionnaire at baseline and endline or just endline?).

Response:

Regarding the secondary outcomes, we have explained above to indicate that the secondary outcome is to understand physician’s barriers to the on-line video-based training programme by conducting a survey on physicians’ education background, working experience, and perceptions, attitudes and practices about the training programme.

We have developed the following data collection and analysis plans as a supplementary file (Supplement 7) of the revised manuscript.

What about harm related outcomes? Are these not necessary? You should justify why if not.

Response:

The intervention group will have an extremely low risk of adverse effects on patients. The potential risk could be disturbance due to regularly reminding of WeChat by project coordinator. We will investigate this potential harm in the post-intervention questionnaire survey.

Data collection

1) When is the physician questionnaire data collected? Baseline and evaluation or just basline?

2) The data collection periods are not clear in relation to the baseline and evaluation surveys mentioned. It seems to imply there is just one data collection period from one month after implementation until at least 6 months after, so where does the baseline survey occur? Is the baseline survey just the "baseline data of prescribing adherence will be retrospectively collected from the most recent 30 archived prescriptions prior to the trial in each of the departments in each of participating hospitals using"? I think it could be explained more explicitly, e.g. "we will collect all outcomes and covariates at baseline and endline, with the baseline survey being... and the endline survey being ..." Also it could be clearer if the primary and/or second outcome data is collected at baseline or is this just the secondary data?

3) You say the data will be "conducted monthly thereafter for at least 6 months". This is not advisable because it allows subjective decision making about how long data are collected for, and therefore opens the door for researchers to influence data collection processes and trial results. This could be criticised later on as questioning the robustness of the processes and results. The data collection period should be fixed and based on good information about the time needed to collect the necessary number of prescriptions.

3) Where does the prescription data come from? Electronic routine records, or paper prescriptions?

4) Who screens the prescriptions to check they are eligible? This must be clear and how independent decisions are maintained, especially around masking? i.e. will screening individuals be blinded or not?

5) What about covariate data, e.g. will you collect data from prescriptions on age, sex etc? List all variables that will be collected. Do the same for any covariates collect at the physician level and/or hospital level, e.g. Dr age, sex, experience/level, hospital size? Or refer to where these data are collected if from the tools in the appendix.

Response:

1) We will have a baseline survey to collect background information of the participating hospitals (Supplement 2) and another baseline questionnaire survey to collect background information of the participating physicians (Supplement 3) in both intervention and control groups for evaluating the appropriateness of randomization. A specific questionnaire survey (Supplement 4) will be conducted among the physicians in the intervention hospitals after implementing the intervention to collect data for measuring the secondary outcome. We made an accordingly update in the revised manuscript to read as “Prior to implementation of the intervention, we will have a baseline survey to collect background information of the participating hospitals (see Supplement 2) and another baseline questionnaire survey to collect background information of the participating physicians (see Supplement 3) in both intervention and control groups for evaluating the appropriateness of randomization. A specific questionnaire survey (see Supplement 4) will be conducted among the physicians in the intervention hospitals after implementing the intervention to collect data for measuring the secondary outcome.” (Page 11)

2) The baseline questionnaires will be used for evaluation of appropriateness of randomization of hospitals and difference between intervention and control groups in participating physician’s characteristics. Audits of prescriptions prior to intervention implementation and 6 months after the implementation will be used to measure the primary outcome. A specific questionnaire survey among participating physicians in the intervention hospitals will be used to measure the secondary outcome. We have clarified these in the relevant parts of the revised manuscript. (Page 11)

3) We have responded to the similar comment above.

3) A clarification of “Based on the Hospital Information System (HIS)” has been made to the revised manuscript (Page 11).

4) Our investigators (prescription assessors) will screen the prescription to obtain the most recent 100 eligible prescriptions from each participating hospital. In the study, the prescription assessors and data analyst will not be aware of the assignments. These issues have been dealt with in responses to the above comments. (Page 11)

5) We have listed the information we will collect from prescriptions (see Supplement 5 and Supplement 6), physicians (see Supplement 3) and hospitals (see Supplement 2).

Data analysis

There is a lot of detail missing currently that will open your results up to strong criticism. This can be added here or included in a separate statistical analysis plan. See comments below for details.

1) You appear to suggest you will test for baseline differences in characteristics with null hypothesis significance testing (chi-sq, t-test and Mann Whitney U specifically). This is not recommended and explicitly discouraged by CONSORT and more generally as the null hypothesis is known to be true (because the groups only differ due to random sampling as they were randomly assigned), so the p-value is completely meaningless. You only need to present descriptive statistics of characteristics for each group, i.e. frequencies and percentages, means and SDs or medians and IQRs etc.

2)You must state clearly the priority of your analyses in terms of which will be used as the primary analysis to determine effectiveness, i.e. the crude or covariate adjusted analysis? As with the secondary outcomes if you do not specify which of these analyses will be used to make the determination about treatment effectiveness then you are leaving open the possibility that you can choose between crude or adjusted analyses as to which gives you the most favourable result. I would suggest your adjusted results are used with the crude results acting simply as a sensitivity analysis to illustrate whether covariate adjustment is important for the results or not. Adjusting increases power and reduces the effect of imbalances between arms with typically minimal risks.

3)You need to be clear about which outcomes you are describing the analysis of (explicitly state), and what effect measures you will estimate, e.g. if the primary outcome then as a proportion it will be a risk difference or risk ratio depending on which approach you take, but this should be clear and isn’t currently. The p-value is not the key measure of effect, that will be the size of effect and 95% CI around that estimate (e.g. risk difference). Similarly, for your secondary outcomes all analyses should be described in the same level of full detail. You also must fully specify which covariates will be adjusted for in your adjusted analyses, and what units and categories those covariates have (e.g. age [years], education [none, primary, secondary, further] etc). This is particularly important in terms of being clear whether the baseline outcome data will be used as a covariate. Again not doing this opens your adjusted analyses up to criticism that you chose covariates that gave you desirable results after collecting the data.

4)You have not explained how you will deal with missing data. This needs a full treatment. E.g. will you use any form of imputation (if so provide full details), or just complete case analysis, and if so what are the justifications for doing so?

5)You do not specify any subgroup analyses. If you do not plan to do any this should be explicitly stated. If you do plan to do any they should be fully detailed (all methods) here. Again if you do not do this then you cannot call them pre-planned in your results paper and they will rightly be viewed as very open to bias (even more than subgroup analyses already are) because again it leaves open the possibility that you selected analyses that suited what results you wanted to find.

6)CONSORT recommends avoiding using the term “intention to treat” because it has no universally accepted definition. Instead you should explicitly describe who will be in your main and any sensitivity analyses. It is particularly important to consider how any missing data and how you deal with that will affect these analysis populations. E.g. if you only do complete case analyses and have some missing data then your analysis populations will not be the full set of originally “recruited” patients whose data you collected. Therefore, it should be completely clear who will be included in the analyses, and any sensitivity analyses (e.g. “per protocol” analyses, which again is a fairly meaningless term and should instead be fully defined in terms of who what will include). Refer to your trial flow chart to help readers understand this better. However, the flow chart needs to be clear whether the internal pilot sites are only intervention sites and if the data collected in the control arm started after the pilot. Consider the relative placement of the boxes as the control arm one is in between the intervention pilot and main trial box which is ambiguous.

7)You say single and multiple regression analyses will be further conducted to identify factors related to nonadherence to the National Guidelines among participants in the intervention arm. This may be better for a subsequent paper than your main trial results paper, but that is of course up to you, however these methods are not well defined again as they are just a few words. First, note that only multi-level regression methods would be suitable given the clustered nature of the data, second, what covariates will be used, third, how will you deal with missing data etc (you need to provide the same details for all analyses). Lastly, it’s unclear why you would do single and multiple regression analyses, given that single regression analyses will just give you more confounded results than multiple regression analyses?

8) State if there are any interim analyses planned or not, and if so provide full details.

Response:

We have prepared the data collection and analysis plans (Supplement 6) to respond to the comments 1) – 6).

6) We have replaced the “intention to treat” with “complete case analyses”, and explicitly described that we will conduct complete case analyses of outcomes unless there is an indication that data may be missing not at random, in which multilevel multiple imputation methods to deal with missingness will be employed as sensitivity analyses in addition to the complete case analyses. (Page 11) We have also updated our trial flow chart (Figure 1) as suggested.

7) We have deleted the single and multiple regression analyses from our protocol as suggested.

8) We have state that no interim analyses are planned in the revised manuscript (Page 12)

Ethics and dissemination: I'd move the consent process details into the methods as it's easier to understand there when explained as part of the recruitment and randomisation process, also it needs to be clear who consents for hospitals to be part of the study, and it should be clear if patients are not asked for consent and if so this should be justified. It should also be clear who recruits hospitals and physicians into the study, e.g. research team members? Again I would say all this detail should go into the methods under an appropriate section (e.g. "recruitment and randomisation" would make sense because then you can explain how these things happen chronologically).

Response:

We have moved the consent process into the methods under the “Participants, recruitment and consent” section as suggested and explained how the recruitment happen chronologically. The statement has changed to “Our investigators will prepare a list of eligible hospitals in each of the 8

study provinces. These hospitals will be invited to participate in the study by explaining the objectives and processes of the study to the hospitals' administrators and the project coordinators. If the hospital agree to participate in the study, the project coordinator will work together with the study investigators to invite physicians to participate in the study. We have added this statement into the revised manuscript. (Page 8) The informed consent process has been described in the previous version of the manuscript.

Reviewer: 2

Reviewer Name: Sun Qiang

Institution and Country: School of Healthcare Management, Shandong University, P.R. China

The manuscript plans to use a cluster randomized controlled trial to evaluate the effects of an education-based intervention on improving the rational use of antibiotics. the design of cluster randomized controlled trail is clearly described and the process evaluation was also considered, the outcome measurements are acceptable. But the key problems of the manuscript is the intervention itself. How can the authors justify the novelty of the education-based intervention? What are the biggest advantages of this video learning materials compared to the traditional printed training booklets? How do the study ensure the participates to look at the video as required? Please justify them. In addition, why the study randomly select 17 prescriptions per months in each hospital? how was the sample size calculated.

Response:

The biggest advantage of the intervention via a video-based learning tool is ease-to-use (user-friendly), possibility of playback (training-continuous), autonomy of learning (time-flexible).

Studies carried out within other medical fields have indicated that on-line video-based education intervention is helpful in guiding appropriate prescribing behaviors, but there are no studies that shed light on whether such training programme is beneficial to improving antibiotic prescribing behaviors among physicians who provide treatment to patients with sexually transmitted diseases (STDs). (Page 6) The current study is aimed to validate this similar impact of this learning tool in these providers.

The visiting frequency of the on-line training video from each intervention hospital will be recorded in real-time on the WenJuanXing platform, therefore we can evaluate whether or not the intervention is delivered as intended from the server logs and will keep in close contact with our in-hospital project coordinators to promote adherence to the intervention ("The intervention of video-based training will last for 6 months and during the 6 months the participating physicians will be reminded for video playback by the project coordinators through sending a WeChat message monthly." (Page 9)).

VERSION 2 – REVIEW

REVIEWER	Joe Hicks Nuffield Centre for International Health and Development, University of Leeds
REVIEW RETURNED	24-Apr-2020

GENERAL COMMENTS	Thank you for addressing all my comments with carefully considered

changes or carefully considered counter-arguments. When you spend a few hours constructively criticising a manuscript it's so frustrating when the authors just ignore the comments they don't like and don't even bother to address them, so it's great to feel like you are really making a positive difference with the peer-review process (as sometimes this doesn't feel to be the case). I think the changes have really strengthened the protocol substantially and make it more robust, and they will make your results reporting much "safer" for you, because there will be far less room for reviewers to (fairly) criticise the robustness of your reporting when you give yourselves less room to modify your design and analyses etc. Of course should the need arise you may have to change certain things, but then you can explain why and point to the original plan in the protocol as evidence that it was a "legitimate" change and not made to "improve" the results.

Unless I have said otherwise I accept the changes you have made or the counter-arguments. I have said my recommendation is to accept the manuscript, but please see below for a few minor counter-comments, and a few additional comments where I have noticed something I feel could be improved that I didn't see before, which I think are well worth addressing as they shouldn't take long and just help make a few minor bits more clear to readers.

As I said previously this is a very interesting and large-scale/powerful study, and as I work with Chinese colleagues on reducing inappropriate antibiotic prescribing in Chinese primary care through similar interventions (for respiratory infections, see Prof. Xiaolin Wei's work if you are interested) this is a study that I know I and they will be very interested to hear the results of. All the best with the work.

Final comments

Study design and setting and study sites

For completeness (e.g. see SPIRIT checklist) add what trial framework this will follow (and unless I've completely misunderstood this will be a superiority trial, i.e. testing whether the intervention is superior/better than the control [rather than equivalent or not inferior), and what the trial design will be (again unless I've misunderstood completely my reading is that it is a "parallel arm"/"parallel group" [same thing] cRCT design), e.g. "a cluster randomised, parallel group, superiority trial". No need to call it "prospective" because you can't have a retrospective trial, only retrospective cohort/observational studies.

You state "Within each study province, we will invite the city-level hospitals with the high volume of reported gonorrhea cases to participate in the study.", and further down in the "Cluster randomization" section you similarly state "Within each province, we will stratify the hospitals (clusters) into two groups (provincial capital

	city and other cities), and 4 hospitals from capital city and 14 hospital from other cities will be purposively selected in consideration of a high volume of reported cases." However, this is still a bit unclear because a "high volume of reported cases" doesn't clarify how high is high, or how low means you don't get selected. e.g. you are selecting 18 hospitals per province. So out of the possible n hospitals per province are these 18 the hospitals (stratified by provincial city and other city) that have the HIGHEST volume of reported cases, or what? It just needs to be clearer what the criteria/threshold is for selecting the 18 vs all other possible hospitals, and it might just be as simple as saying "highest" instead of "high", as that makes it clear there is a threshold and what it is, because "high" is a little vague. Fig 2 is helpful but you state "The 8 provinces (Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong, Guangxi, Hainan and Yunnan) are purposively selected from 31 provinces in mainland China in consideration of their higher incidence of reported gonorrhea cases than the national average", so I wonder do you mean they are the 8 provinces with the highest prevalence out of the 31 mainland provinces? If so it would be clearer to state this in these terms, i.e. "The 8 provinces (Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong, Guangxi, Hainan and Yunnan) are purposively selected from the 31 provinces in mainland China due to having the highest incidence of reported gonorrhea cases." Otherwise it's a bit unclear because your phrasing leaves open the possibility that there are >8 provinces with an incidence higher than the national average, but based on some unstated criteria/reasoning you just selected 8 (which may not be the 8 highest). I hope that is clear. You state: "Within each study province, we will invite the city-level hospitals with the high volume of reported gonorrhea cases to participate in the study." Please be more specific, as it's not clear what this means. e.g. will you select a threshold that applies to all provinces, or select different thresholds within each province, and how will they be chosen, on what basis etc? I think you have misunderstood my comment regarding selection of the pilot province. I did not (and would not) specify that you should select the pilot province randomly, that is your decision and either randomly or purposive approaches could be justifiable. My comment was just asking for a bit more clarity about how you will select the province, whether randomly or purposively. So if you prefer a purposive selection then stick with that, but just explain briefly why and what criteria/characteristics will be used to make that purposive selection. e.g. you might choose the province with the highest prevalence to test if it can be feasible in an "extreme" situation, or maybe you might choose a province (out of the 8) with the median prevalence to see how it works with a "typical" prevalence etc. Just be more specific and detailed (within reason) so it's clear what the rationale is.

	You state that the heterogeneities between provinces can help with representativemess, and I completely agree. The study scale is extremely large and helpful for generalisability, but my comment asking you to explain this heterogeneity in more detail and highlight it explicitly so readers can understand how different the provinces are, and therefore how representative the findings are likely to be. # Outcomes I have read the supplement 7 and you provide enough info to reasonably clearly define a "non-adherence" outcome for Drs, and say "We will use the nonadherence as dependent variables for the analyses to identify the factors influencing the nonadherence". But the wording in the main protocol is a bit confusing because you say "The secondary outcome is to understand physician's barriers to the on-line video-based training programme, including physician's socio-demographic characteristics, education ...". I think you should simply define the secondary outcome, which appears to be Dr adherence <25%, and leave it at that. The outcomes section of a protocol is not where you need to explain any aims of the analyses, or analysis objectives/processes, so just define the outcomes. You can explain how you will analyse them and what factors you will look at (beyond the treatment arm) in the statistical analysis section/statistical analysis plan. Make sure all outcomes you plan to analyse are fully defined, or you as I previously discussed you will have problems reporting them because they could be criticised as unplanned. See the SPIRIT checklist guidance on reporting outcomes (the CONSORT checklist/guidance for cRCTs is very similar): "Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended" The info in the supp 7 covers these criteria, but for the outcomes section of the main protocol you could state it more concisely than is done there. # Statistical analysis Your statistical analysis plan supplementary material is much more comprehensive now. However, my advice for the secondary outcome analysis would be to simply select the covariates in your model based on theory/expert judgement, and not use step-wise selection processes. I know they are extremely popular and still taught widely, but there is plenty of evidence and theory that they lead to a biased type 1/false-positive error rate, because by definition you are letting the data select covariates that explain variation in the outcome, so you make it more likely (in a false biased way) that you will get a significant result. A quick Google will

	bring up plenty of article/blogs etc explaining the substantial problems with step-wise processes (outside of prediction modelling that is). An unintended side effect of the theory driven approach is also you reduce your workload a lot because you aren't having to run loads of model selection analyses.

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Joe Hicks

Institution and Country: Nuffield Centre for International Health and Development, University of Leeds

Please leave your comments for the authors below

Thank you for addressing all my comments with carefully considered changes or carefully considered counter-arguments. When you spend a few hours constructively criticising a manuscript it's so frustrating when the authors just ignore the comments they don't like and don't even bother to address them, so it's great to feel like you are really making a positive difference with the peer-review process (as sometimes this doesn't feel to be the case). I think the changes have really strengthened the protocol substantially and make it more robust, and they will make your results reporting much "safer" for you, because there will be far less room for reviewers to (fairly) criticise the robustness of your reporting when you give yourselves less room to modify your design and analyses etc. Of course should the need arise you may have to change certain things, but then you can explain why and point to the original plan in the protocol as evidence that it was a "legitimate" change and not made to "improve" the results.

Unless I have said otherwise I accept the changes you have made or the counter-arguments. I have said my recommendation is to accept the manuscript, but please see below for a few minor counter-comments, and a few additional comments where I have noticed something I feel could be improved that I didn't see before, which I think are well worth addressing as they shouldn't take long and just help make a few minor bits more clear to readers.

As I said previously this is a very interesting and large-scale/powerful study, and as I work with Chinese colleagues on reducing inappropriate antibiotic prescribing in Chinese primary care through similar interventions (for respiratory infections, see Prof. Xiaolin Wei's work if you are interested) this is a study that I know I and they will be very interested to hear the results of. All the best with the work.

Response:

We appreciate the reviewer’s positive comments. As suggested, we have addressed the minor comments and made updates accordingly in the revised manuscript.

Final comments

Study design and setting and study sites

For completeness (e.g. see SPIRIT checklist) add what trial framework this will follow (and unless I've completely misunderstood this will be a superiority trial, i.e. testing whether the intervention is superior/better than the control [rather than equivalent or not inferior), and what the trial design will be (again unless I've misunderstood completely my reading is that it is a "parallel arm"/"parallel group" [same thing] cRCT design), e.g. "a cluster randomised, parallel group, superiority trial". No need to call it "prospective" because you can't have a retrospective trial, only retrospective cohort/observational studies.

Response:

We have added the trial framework and trial design in the revised manuscript to read as “We will use a parallel-group, cluster randomized controlled trial with one intervention and one control arm, using a 1:1 allocation ratio, to evaluate whether the intervention is superior to the control treatment (Figure 1).” (Page 5)

We have deleted the word “prospective” in the revised manuscript as suggested.

You state "Within each study province, we will invite the city-level hospitals with the high volume of reported gonorrhea cases to participate in the study.", and further down in the "Cluster randomization" section you similarly state "Within each province, we will stratify the hospitals (clusters) into two groups (provincial capital city and other cities), and 4 hospitals from capital city and 14 hospital from other cities will be purposively selected in consideration of a high volume of reported cases." However, this is still a bit unclear because a "high volume of reported cases" doesn't clarify how high is high, or how low means you don't get selected. e.g. you are selecting 18 hospitals per province. So out of the possible n hospitals per province are these 18 the hospitals (stratified by provincial city and other city) that have the HIGHEST volume of reported cases, or what? It just needs to be clearer what the criteria/threshold is for selecting the 18 vs all other possible hospitals, and it might just be as simple as saying "highest" instead of "high", as that makes it clear there is a threshold and what it is, because "high" is a little vague.

Response:

Sorry for the unclear description. We have replaced the word “high” with “highest” to explicitly clarify how the 18 hospitals will be selected. (Page 6, Page 9)

Fig 2 is helpful but you state "The 8 provinces (Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong, Guangxi, Hainan and Yunnan) are purposively selected from 31 provinces in mainland China in consideration of their higher incidence of reported gonorrhea cases than the national average", so I wonder do you mean they are the 8 provinces with the highest prevalence out of the 31 mainland provinces? If so it would be clearer to state this in these terms, i.e. "The 8 provinces (Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong, Guangxi, Hainan and Yunnan) are purposively selected from the 31 provinces in mainland China due to having the highest incidence of reported gonorrhea cases." Otherwise it's a bit unclear because your phrasing leaves open the possibility that there are >8 provinces with an incidence higher than the national average, but based on some unstated criteria/reasoning you just selected 8 (which may not be the 8 highest). I hope that is clear.

Response:

Sorry for the unclear description. As suggested, we have corrected it to read as “The 8 provinces (Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong, Guangxi, Hainan and Yunnan) are purposively

selected from the 31 provinces in mainland China due to having the highest incidence of reported gonorrhea cases (Figure 2)". (Page 6)

You state: "Within each study province, we will invite the city-level hospitals with the high volume of reported gonorrhea cases to participate in the study." Please be more specific, as it's not clear what this means. e.g. will you select a threshold that applies to all provinces, or select different thresholds within each province, and how will they be chosen, on what basis etc?

Response:

Similar to our response to the above comment on hospital selection criteria, we have made it clear to read as "Within each study province, we will invite the city-level hospitals with the highest volume of reported gonorrhea cases to participate in the study." (Page 6)

I think you have misunderstood my comment regarding selection of the pilot province. I did not (and would not) specify that you should select the pilot province randomly, that is your decision and either randomly or purposive approaches could be justifiable. My comment was just asking for a bit more clarity about how you will select the province, whether randomly or purposively. So if you prefer a purposive selection then stick with that, but just explain briefly why and what criteria/characteristics will be used to make that purposive selection. e.g. you might choose the province with the highest prevalence to test if it can be feasible in an "extreme" situation, or maybe you might choose a province (out of the 8) with the median prevalence to see how it works with a "typical" prevalence etc. Just be more specific and detailed (within reason) so it's clear what the rationale is.

Response:

Thanks for the reviewer's explanation. We decided to select the pilot province randomly.

You state that the heterogeneities between provinces can help with representativeness, and I completely agree. The study scale is extremely large and helpful for generalisability, but my comment asking you to explain this heterogeneity in more detail and highlight it explicitly so readers can understand how different the provinces are, and therefore how representative the findings are likely to be.

Response:

Thanks for the review's further explanation. We have explained the heterogeneity in the revised manuscript to read as "According to geographic location and the level of economic development, the mainland China was classified into three areas -- Eastern areas (high level of economic development), Middle areas (middle level of economic development) and Western areas (low level of economic development). Among the 8 selected provinces, Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong and Hainan are in the eastern areas, while Guangxi and Yunnan are in the western areas. In China, health workers (especially highly educated physicians) prefer to serve in economically developed areas rather than poverty-stricken areas, which results in a severe inequality in quantity and quality of health workforces. The training of the health workers is considered as an important way for enhancing professional skills, but the effects of training may vary from area to area. Therefore, by evaluating the intervention effects in Chinese hospitals in multiple provinces representing different economic levels and geographic regions, our study may provide potentially

useful information about the extent to which implementation of the intervention is likely to be influenced by local context." (Page 6)

Outcomes

I have read the supplement 7 and you provide enough info to reasonably clearly define a "non-adherence" outcome for Drs, and say "We will use the nonadherence as dependent variables for the analyses to identify the factors influencing the nonadherence". But the wording in the main protocol is a bit confusing because you say "The secondary outcome is to understand physician's barriers to the on-line video-based training programme, including physician's socio-demographic characteristics, education ...". I think you should simply define the secondary outcome, which appears to be Dr adherence <25%, and leave it at that. The outcomes section of a protocol is not where you need to explain any aims of the analyses, or analysis objectives/processes, so just define the outcomes. You can explain how you will analyse them and what factors you will look at (beyond the treatment arm) in the statistical analysis section/statistical analysis plan. Make sure all outcomes you plan to analyse are fully defined, or you as I previously discussed you will have problems reporting them because they could be criticised as unplanned. See the SPIRIT checklist guidance on reporting outcomes (the CONSORT checklist/guidance for cRCTs is very similar):

"Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended"

The info in the supp 7 covers these criteria, but for the outcomes section of the main protocol you could state it more concisely than is done there.

Response:

Thanks for this valuable suggestion. The outcomes section has been updated in the main protocol to read as "The primary outcome is the proportion of prescriptions adherent to the regimens for treatment of uncomplicated gonorrhea recommended by the National Guidelines at the cluster level, which will be documented at baseline and final evaluation. The second outcome is a physician-level binary indicator of nonadherence (defined as having at least 25% of prescriptions nonadherent to the National Guidelines)." (Page 8-9)

Statistical analysis

Your statistical analysis plan supplementary material is much more comprehensive now. However, my advice for the secondary outcome analysis would be to simply select the covariates in your model based on theory/expert judgement, and not use step-wise selection processes. I know they are extremely popular and still taught widely, but there is plenty of evidence and theory that they lead to a biased type 1/false-positive error rate, because by definition you are letting the data select covariates that explain variation in the outcome, so you make it more likely (in a false biased way) that you will get a significant result. A quick Google will bring up plenty of article/blogs etc explaining the substantial problems with step-wise processes (outside of prediction modelling that is). An unintended side effect of the theory driven approach is also you reduce your workload a lot because you aren't having to run loads of model selection analyses.

Response:

Thanks for the reviewer's advice. After referring to the relevant articles, we decided to select the covariates in our model based on theory, and we have updated the description in the Supplement 7 to read as "We will use multivariate logistic regression model to explore the associations between key factors (physician's age [years], education level [bachelor degree or higher, other], working year [years], department [dermatology or dermatovenerology, other], on-line training video access times [≤3,>3], perceptions on the National Guidelines [appropriate (enough) dosage, other], attitude on the training programme [very helpful or helpful, other]) and the nonadherence. These results will be presented as adjusted odds ratio (AOR), associated 95% CIs and p values. The factors included in the model will be determined theoretically, based on the literature and the findings of our qualitative work." (Supplement 7)

VERSION 3 - REVIEW

REVIEWER	Joe Hicks Nuffield Centre for International Health and Development, University of Leeds
REVIEW RETURNED	12-May-2020

GENERAL COMMENTS	Dear authors, I thought this was ready for acceptance previously (and I indicated this) and my minor comments were optional, so I assume it will be accepted now. All the best with this important research.