PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The STOP-AB Trial protocol: Efficacy and safety of discontinuing patient antibiotic treatment when physicians no longer consider it necessary.
AUTHORS	Llor, Carl; Moragas, Ana; Bayona, Carolina; Cots, Josep M.; Molero, José M.; Ribas, Joana; Fóthy, Julio Francisco; Gutiérrez, Isabel; Sánchez, Coro; Ortega, Jesús; Arranz, Javier; Botanes, Jenifer; Robles, Purificación

VERSION 1 - REVIEW

REVIEWER	Gloria Cordoba Department of General Practice University of Copenhagen Denmark I know personally two of the authors.
REVIEW RETURNED	16-Jan-2017

GENERAL COMMENTS	SUMMARY This RCT aims to compare two management strategies (Discontinuation of antibiotics vs continuation of antibiotics) in adult patients (18-75 year old) with an uncomplicated respiratory tract infection (RTI). The main outcome is number of days with severe symptoms and the sample size calculation is based on an assumption of one day difference of the number of days with severe symptoms between the two groups. The topic of the study is highly relevant not only within the Spanish setting, but also worldwide. Discontinuing an unnecessary antibiotic treatment is a frequent dilemma faced by clinicians, but there is limited knowledge about the safety of implementing this practice, thus it is difficult to convince clinicians that discontinuation of antibiotics is an option they should discuss with their patients. MAJOR REVISION: 1. In the abstract at the methods and analysis section (line32-36 page 3) you mention other two study designs, while in the main manuscript (from line 3 to 14 in page 14), you write about these two post-trail designs in the discussion section. Then, you have to decide: a) If the main objective of this protocol is to focus only on the methodology of the RCT, you have to delete (line32-36 page 3) b) If the main objective of this protocol is to describe the

	methodology of the whole study that includes 3 different types of study designs, then you have to expand the description of the observational and qualitative study in the methodology section. 2. You need to explain how you will secure that GPs do not vary so much in their assessment of uncomplicated RTI. It is important because you are assuming quite a homogeneous judgement as far as you are not planning to control by the effect of clustering (e.g. existence of common guidelines, intensity and content of the training prior to data collection). 3. The main and secondary outcomes will be measure with questionnaires. You should explain more about the validity of these questionnaires for the population in which they will be used and whether there is any limitation/weakness arisen from using these questionnaires. 4. Regarding the criteria for withdrawals, please indicate: who will judge the “unsatisfactory therapeutic effect”? the doctor or the patient? (line 38 pag 12) How will you decide that there is an “unsatisfactory therapeutic effect”? 5. The discussion part would benefit from a detailed account of the strengths and weaknesses of the methodology, data collection instruments and analysis. It would be good as well if you highlight the expected impact of confirming the hypothesis of the study. MINOR REVISION: 1. Page 4. Following the title, please write first the strengths and afterwards the limitations. Check that everything you wrote in this page is part of the discussion section 2. Line 41 page 8 “an effect on their beliefs is expected”. The beliefs of the doctors or the patients…. Or actually both? 3. Page 5 line 55. Grammar issue: who is the subject of the action? The GP? Or the patient? 4. I would suggest language revision by a native English speaker

REVIEWER	Lisa Bebell Massachusetts General Hospital and Harvard Medical School, USA
REVIEW RETURNED	22-Jan-2017

GENERAL COMMENTS	This protocol is designed to determine the safety of antibiotic discontinuation among outpatients originally prescribed antibiotics, but suspected to have viral diseases. The main hypothesis put forth by the authors is that although antibiotic discontinuation is recommended for suspected viral infections, there is no evidence that stopping antibiotics is safe. The authors argue that the Spanish Society of Family Medicine has been recommending GPs to ask their patients to stop taking antibiotics when they suspect a viral infection since 2011, but this recommendation is not evidence-based. Thus, they have designed a multicenter clinical trial protocol to determine the safety of this recommendation in adult patients. The protocol is generally well-written. The subject is of broad interest

	to clinicians practicing in nearly all settings, as this clinical situation is commonly encountered regardless of geographic location and resource availability. The main hypothesis is reasonable, though the extensive and vague exclusion criteria and high proportion of projected loss to follow-up concern me that the results will not be generalizable in a meaningful way. I believe the paper would be substantially improved by increasing the specificity of the inclusion and exclusion criteria, decreasing the number of clinicians and centers involved, and providing the tools to be used for qualitative data collection. Major and minor points are described in detail below. Major Points: 1. Patients must consent to participate in this open-label, non-blinded study. I am concerned that the population agreeing to participate will not be representative of the Spanish outpatient population at large, and therefore that results of this trial may not be generalizable enough to be useful in clinical practice. Specifically, patients (and clinicians) agreeing to participate in this trial are likely to be more open to the idea of stopping antibiotics in the case of suspected viral illness. Trial participants may be less likely to have taken non-prescribed antibiotics, or may be untruthful about their antibiotic use in a way not representative of your outpatient population. This could also introduce information bias and lead to a misinterpretation of the results. 2. How will the authors account for variability in experience, confidence and comfort levels of GPs in their assessment that “the GP considers that antibiotics are not needed to be taken”? It seems possible that some GPs in your study could have abnormally low or high confidence in their ability to assess whether antibiotics are necessary, this could skew the results significantly, and potentially lead to a greater number of adverse events. Perhaps inviting GPs who have a midrange level of experience (e.g. have been in practice for 5-15 years) might help to equalize their confidence levels? Another strategy would be to administer a questionnaire to the GPs using vignettes or questions to assess their level of confidence in stopping antibiotics, which might make your results more interpretable. 3. The authors propose secondary study objectives to study the incidence adverse events (page 9, objective 1) and number of complications within the first 3 months (page 9, objective 4). Though these are not primary objectives, I believe the study will be underpowered to detect most adverse events and complications, and this limitation should be recognized more clearly. 4. The authors propose additional qualitative secondary study objectives to assess antibiotic consumption (page 9, objective 2) and assess satisfaction (page 9, objective 3). However, the tools used for these assessments are not provided, and the authors do not adequately acknowledge the potential for information bias when relying on patient self-report of these measures. 5. I am concerned that the authors predict a 20% proportion of enrollees will be lost to follow-up (page 10, lines 53-55). This is a high proportion lost, and these patients are likely to be significantly different to those retained in the study, including their compliance

	with antibiotic recommendations. Loss of this high proportion of participants could significantly affect trial results. 6. The authors plan to invite 40 clinicians from 20 different healthcare settings to join their study, anticipating that each clinician will recruit 10-12 participants over the 2-year study period to achieve their projected enrollment of 430 total participants. Though it is a noble goal to include clinicians and patients from diverse practice settings, I am concerned that it will not be feasible or practical to ensure that 40 clinicians can be trained and achieve compliance with the study protocol and recommendations, and retain these skills over a 2-year period while enrolling participants infrequently. Moreover, recruitment of 5-6 participants annually (1 every 2 months) represents a very small proportion of all patients seen in each provider's clinical practice. If, as the authors state, 15% of all patient visits are for viral conditions targeted by this trial, I would estimate that most clinicians would see approximately 1-3 patients every day with viral illnesses that might qualify them for this study. The expectation that each clinician would recruit such a small proportion of the eligible patient population concerns me that clinicians would 'cherry-pick' patients they wished to enroll, choosing to approach potential research subjects not at random, but rather in a way that is convenient to their clinical practice. In my opinion, this has great potential to introduce systematic bias into the sample, and invalidate the results. I think the trial would be more valid if it were restricted to a smaller number of clinicians (~10), and these clinicians were mandated to approach all eligible patients sequentially for enrollment, until they reached their target enrollment of 43 participants each. In addition, I would request the authors specify what data can be collected from participants who chose not to enroll, important information to help future readers decide how applicable trial results are to their own population. Minor Points: Page 5 1. Lines 31-32: I think it is an overstatement to state "General practitioners (GP) have always been told to continue an antibiotic regimen once the patient has initiated it in order to prevent the patient from acquiring resistant organisms." Though I am unfamiliar with general practice in Spain, using the word always might be too strong. I recommend softening this statement to eliminate that word and instead state "Since 2011, general practitioners (GPs) in Spain have been told to continue . . ." 2. Lines 34-52: This section discusses the appropriate dosing of antibiotics for treatment of diagnosed bacterial infections. While the authors make valid points, I think this discussion is tangential to the focus of the protocol and could be eliminated or significantly shortened so as not to distract from the main focus of the paper, which is about antibiotic use in non-bacterial infections. 3. Lines 54-56: Perhaps you are describing a local Spanish practice of making home visits to patients, but if you intend to describe office visits with patients, I recommend re-wording this sentence to say "GPs often see patients with suspected viral infections of the upper and lower airways, for which antibiotic

	treatment makes no difference in terms of clinical outcomes." Furthermore, to state that antibiotic treatment makes no difference is a strong statement. Some antibiotics have known anti-inflammatory properties and may improve the outcome of viral illness by this mechanism. In addition, to my knowledge there is not enough evidence to support the statement, and in fact, a negative impact (e.g. an antibiotic-related side effect or new case of C. difficile diarrhea) could be viewed as a difference in clinical outcome. Please consider rewording this statement with this view in mind. Page 6 1. Lines 4-5: Consider exchanging 'bugs' for 'bacteria'. 2. Lines 6-7: Why should this be common practice? I follow your argument, but there are many reasons why it should not be common practice, including many behavioral aspects of clinical medicine including the desire to give some type of treatment to the patient, patient demand of medications, etc. I recommend that you consider rewording this to take into account opposing views. 3. Lines 11-13: This statement about GPs feeling unsafe to discontinue an antibiotic seems reasonable, but you do not support it with a citation. Can you provide a reference for this assertion? 4. Lines 32-34: See my comments above about 'no difference.' Perhaps this could be reworded to state that antibiotics 'do not improve outcomes' for viral upper respiratory infections? 5. Lines 25-27: A proportion of 2.2% of all episodes of common cold receiving antibiotic treatment does not seem high to me, and is what I would expect under the very best clinical practice. Is this a typographical error? 6. Lines 49-56: This is an overly long heading and slightly confusing. Can it be shortened and reworded to state the problem more clearly? Page 7 1. Lines 3-22: The authors state that antibiotics "might" be required to treat these infections, but that "most" are self-limiting. In my mind, this is the kind of statement that drives antibiotic use, because clinicians are not certain whether they are in the "might" category or the "most" category. I note this as a potential barrier to implementing your trial, and also because it seems to me a rational (rather than irrational) reason for antibiotic use that may be hard to address. It might be helpful to reword this section to address this conundrum, and set up the reader to understand how you will address this particular problem in your trial. Page 9 1. Line 18: Please change 'and' to 'or' for clarity. 2. Lines 32-34: How will you assess for an 'inadequate family setting'? Do you mean homelessness? Or something else? These criteria seem vague and might benefit from more specificity. 3. Lines 45-46: It might help to specify the timeline for a 'terminal disease', e.g. life expectancy less than 6 or 12 months. 4. Lines 49-50: Similarly to above, it is vague what you mean

	by ‘difficulty to attend the programmed visits’. Can you specify a distance from clinic or another measure that would clarify this exclusion? Page 12 1. Lines 19-20: The authors state they will use Fisher’s exact test to compare outcomes in the randomized groups. This test may be difficult to use given the number of proposed participants (430). A Chi2 test may be more appropriate, when estimated cell sizes are >5. 2. Lines 36-39: One of the criteria for withdrawal is “unsatisfactory therapeutic effect”. This is vague and could lead to protocol discontinuation for a number of reasons. Can you be more specific, so the reader can understand how this will not introduce bias? I am concerned that participants in the intervention group will be dissatisfied with ongoing symptoms after being told to stop antibiotics, and will withdraw from the study in greater numbers than those in the control group, affecting your results.

REVIEWER	Gwendolyn (Lyn) Gilbert University of Sydney Australia
REVIEW RETURNED	31-Jan-2017

GENERAL COMMENTS	Most of my "yes" and "no" answers above should be qualified by "partially", "maybe" etc. I hope the following comments clarify them. The paper describes an open label randomised controlled trial of the safety of discontinuing patient antibiotic treatment when the physician does not consider it necessary. It aims to test the common fear among GPs that it is harmful to stop a course of antibiotics once started, even if it was not needed in the first place. Whilst this may be a common belief, it has never been clear why it should be harmful in these circumstances. Presumably the fear is that the judgement, that an antibiotic was not needed, was incorrect. However, the same would apply to the decision not to start antibiotics. The only plausible adverse outcome of stopping an antibiotic prematurely is that symptoms would persist or recur if the decision was wrong – in which case it can be started or restarted for the small proportion of patients affected. There is no evidence that stopping the antibiotic before the course is finished will increase the risk of acquiring a resistant organism – on the contrary, there is evidence that the longer the course the greater the risk. Moreover, the length of a course of antibiotic is generally quite arbitrary and the relatively few studies comparing shorter, with “conventional” courses usually demonstrate equivalence. On the other hand there are many plausible benefits of stopping or not starting unnecessary antibiotic therapy for the majority of patients with viral infections who did not need it. Based on systematic reviews cited by the authors the proportion of patients, in whom symptoms persist for longer when antibiotics were not given for sore throat or acute rhinosinusitis, is significant but very small and the additional duration of symptoms short. Presumably these

	differences would be even less in the group of patients to be studied in this trial, who have already started antibiotics, depending on how long they have been taking them. Two or three days of antibiotics might be as adequate a “course” as a full 5 or 7 day conventional course. These considerations are not well presented in the abstract or background and the manuscript would benefit from some elaboration of these points. On page 6 the authors state that there is no evidence that it is safe to interrupt a course of antibiotics that the GP has judged to be unnecessary. On the other hand there is also no evidence that this practice is, and many plausible reasons why is would not be, unsafe. On page 7, it would be useful to elaborate (with some data from the abstracts of the cited papers) on the statement “recent systematic reviews have suggested……etc.” (lines 3-4). Page 8. Study design. I wonder whether it is right to describe the control group as the one being treated by the “usual strategy”, when the practice of continuing the antibiotic when the GP thinks it is no longer necessary is specifically discouraged by the Spanish Society of Family Medicine? I understand that this is contrary to the other (unfounded) belief that it is dangerous to stop a course, once started and that GPs are understandably confused, but wonder what proportion of “GPs (actually) are reluctant and fell unsafe to discontinue……..” if the GP her/himself believes it is unnecessary and especially if s/he did not prescribe it in the first place. Page 9, lines 3-4. The eligibility criterion “……..patients feel that the antibiotic regimen has not worked as expected and fell they need clinical reassessment” is not clear. If they are randomised to the control group will they be continued on the antibiotic they have already started or changed to another one (and of the latter, will the same antibiotic regimen reused for all patients in the control group in all practices)? For patients in the control group, how will the “course” of antibiotic to be completed be defined? What if they are taking a leftover antibiotic that is inappropriate or out of date,? Will they finish what they have in the cupboard at home or start a fresh course? If the latter, which antibiotic will be prescribed and will they take only as much of the new course to make up the equivalent of a conventional course? Or will they start and complete a new course of the same or a different antibiotic – which will mean a longer than usual total course? If patients in the intervention group have continued symptoms on review – what will be the criteria for recommencing antibiotic therapy and how will they be standardised – number of symptoms, severity? With which antibiotic - the one they had already started initially or another chosen by the GP? Will it be the same for all intervention group patients? Data analysis: page 12. I am not a statistician, but such a relatively small group of subjects and such a large number of different conditions, symptoms and severity gradings might make meaningful analysis difficult. Will severity gradings be equivalent across all different symptoms and conditions?

	What is the definition of a serious adverse event or unsatisfactory therapeutic effect? Do serious adverse events include complications of untreated infections and adverse effects of antibiotic therapy? If patients in control group are able to interrupt medication during the study can those in the intervention group recommence the antibiotic they had started themselves before entry into the study? Finally there are a number of minor editorial/grammatical issues: P 6 line 1. “bugs” replace with “bacteria” p11, lines 21. “change or commence antibiotic treatment” line 28. “…need to commence, change, continue or cease the antibiotic treatment ……..” Presumably a few patients will develop antibiotic side-effects that necessitate stopping prematurely. P13, line 23 “…..withhold the discontinuation of….” Is awkward and confusing suggest “continue” p14, lines 1-3. The sentence beginning “A post-trial implementation….” is confusing – needs rewording.

REVIEWER	Jesse Berlin Johnson & Johnson, USA I'm a full-time employee of Johnson & Johnson. I don't believe there's a direct conflict in reviewing the methodology being proposed in this protocol.
REVIEW RETURNED	15-Feb-2017

GENERAL COMMENTS	1. Throughout the protocol, there are places where the English (grammar and style) could use some attention. 2. I’ll have mostly conceptual questions below. I won’t have a lot of specific detailed suggestions. (Just letting you know.) 3. Abstract: This is more of a general question. I would just ask whether “safety” is the right word to describe your primary objective. Does longer duration of severe symptoms (in terms of number of days) really meet the definition of a “safety” issue? I would think of it more as an efficacy (or reduced efficacy) problem. It’s a semantic point for your consideration, but when I think of safety, from a regulatory/regulated industry perspective, I tend to think more about actual adverse events. These could result from inadequate treatment, so I don’t feel strongly about making any changes. Again, for your consideration. 4. The biggest challenge I had was understanding how you will operationalize the study, including how to define (in practice) the eligible population and how to get clinicians to enroll their patients. You say, “We will include patients from 18 to 75 years of age with uncomplicated acute respiratory tract infections (RTIs) who have previously taken any dose of antibiotic but physicians no longer consider it necessary.” I’m not a clinician, so I may not appreciate or understand what happens in practice (other than having gone to my doctor with a sinus problem in the past). My apologies if I’m missing your point, but here is my problem. You’re saying that the clinician will somehow recognize that a patient is inappropriately receiving an

	antibiotic. Basically the entry criterion is that the clinician says “this person should not have gotten an antibiotic.” For the trial to be successful, the clinician needs to be in equipoise with respect to continuing the antibiotic. That seems to conflict directly with the premise that this same clinician recognizes the inappropriate nature of the original prescription (which he or she may have written). They wouldn’t be considering discontinuation if they didn’t believe it would be inappropriate to continue. Will clinicians really be willing to randomize patients? Would they say, “this patient is inappropriately on an antibiotic but I’m afraid to stop in any case?” Again, maybe I’m misunderstanding. 5. Having said what I just did, it’s possible that a clinician could decide to stop treatment with the antibiotic because of declining symptoms. In that case, I could see that clinician being in equipoise. If that’s the situation you have in mind, maybe it would be helpful to stratify the randomization by reason for stopping (symptoms improved so maybe I can stop vs. condition not improving so maybe it’s a viral infection vs. the culture results came back showing the infection is NOT bacterial?) Maybe there are different strata than these, but I hope you get the idea. I think things could look very different depending on why the clinician thinks it MIGHT be OK to stop. 6. You use duration of symptoms as the outcome measure. Duration from what time point? What is time zero? Is it the time of randomization? Time at which the antibiotic is actually stopped makes sense in the “stop treating” arm but there’s no equivalent (except the time of randomization, I think) in the control (continue treatment) arm. This is a matter of minor clarification, I think. 7. Did you pilot test the symptom diaries? Will subjects be compliant? This diary is the major outcome measure. 8. Secondary objective: You list assessing the incidence of adverse effects of medication as a secondary objective. Technically, that’s only possible in the continued treatment arm, by definition, no? Adverse EVENTS (not necessarily EFFECTS of the drug) could occur in either arm. (The exception, I suppose, would be delayed drug effects.) In any case, I think some additional detail would be helpful here. 9. ON A RELATED POINT: I wasn’t quite clear about your exclusion criterion: “Lack of tolerance to oral treatment, such as the presence of nausea and vomiting, gastrectomy, post-surgery and/or diarrhea.” Are you saying that if someone is having the antibiotic stopped specifically because of an adverse event, i.e., the patient PRESENTS with what is “clearly” (highly likely to be) an adverse effect of the drug, that should disqualify that person from participating in the trial because it would be unethical to continue treatment in that person. That’s fine, and I agree with the rationale (if I’m getting it correct), but again, some clarification will be helpful. As a relevant aside, if you are excluding people who had an adverse EFFECT (here I do mean effect) of the drug, that will reduce the probability that people who are randomized to continued treatment will have a subsequent adverse effect from the drug. 10. Can you define “complications related to the RTI?” How do you distinguish these from other adverse events or symptoms?

	11. I'm assuming the sample size is TOTAL? Not per group? Based on the SD of 3.3? It's a picky technical point, but there's an underlying normality assumption if you're basing your calculation on the t-test (which is what the statistical analysis section indicates). One could argue for some other (e.g., Poisson distribution) approach, as you are counting days, and the Poisson distribution is often used to model counts. It seems very likely that the distribution of duration will be asymmetric (highly skewed). If the mean is 3.6 and SD is 3.3, there is likely a long tail to the right. Similarly – is the t-test the right test for comparing duration?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Gloria Cordoba
Institution and Country: Department of General Practice, University of Copenhagen, Denmark
Please state any competing interests or state 'None declared': I know personally two of the authors.
SUMMARY
This RCT aims to compare two management strategies (Discontinuation of antibiotics vs continuation of antibiotics) in adult patients (18-75 year old) with an uncomplicated respiratory tract infection (RTI). The main outcome is number of days with severe symptoms and the sample size calculation is based on an assumption of one day difference of the number of days with severe symptoms between the two groups.
The topic of the study is highly relevant not only within the Spanish setting, but also worldwide. Discontinuing an unnecessary antibiotic treatment is a frequent dilemma faced by clinicians, but there is limited knowledge about the safety of implementing this practice, thus it is difficult to convince clinicians that discontinuation of antibiotics is an option they should discuss with their patients.
MAJOR REVISION:
1. In the abstract at the methods and analysis section (line32-36 page 3) you mention other two study designs, while in the main manuscript (from line 3 to 14 in page 14), you write about these two post-trail designs in the discussion section. Then, you have to decide:
a) If the main objective of this protocol is to focus only on the methodology of the RCT, you have to delete (line32-36 page 3)
b) If the main objective of this protocol is to describe the methodology of the whole study that includes 3 different types of study designs, then you have to expand the description of the observational and qualitative study in the methodology section.
Response: Thank you for this comment. We have decided to describe only the clinical trial in this protocol. So, we have deleted the last lines of the Methods and Analysis section of the abstract as suggested in option 'a'.

2. You need to explain how you will secure that GPs do not vary so much in their assessment of uncomplicated RTI. It is important because you are assuming quite a homogeneous judgement as far as you are not planning to control by the effect of clustering (e.g. existence of common guidelines, intensity and content of the training prior to data collection).
Response: This a good point and was also suggested by other reviewers. In order to obtain the maximum homogeneity across clinicians, only experienced GPs (those working for more than 15 years) and those who feel comfortable with the design of the study will participate in the study. This will be achieved by administering them a questionnaire with clinical questions, recommendations of guidelines and vignettes and check if they are confident and comfortable with the strategy of stopping an antibiotic course already commenced. Only those who are confident with this strategy will be invited to participate in this clinical trial.

3. The main and secondary outcomes will be measure with questionnaires. You should explain more about the validity of these questionnaires for the population in which they will be used and whether there is any limitation/weakness arisen from using these questionnaires.
Response: We will use validated questionnaires, which have also been used in a previous recently published (reference number 37) randomised clinical trial about the efficacy and safety of delayed prescribing of antibiotics for the same RTIs.

4. Regarding the criteria for withdrawals, please indicate: who will judge the "unsatisfactory therapeutic effect"? the doctor or the patient? (line 38 pag 12) How will you decide that there is an "unsatisfactory therapeutic effect"?
Response: As also suggested by reviewer 3, we have deleted the unsatisfactory therapeutic effect from this paragraph. As already described in the Data collection and ascertainment of visits section, the patients will be interviewed 2-3 days after their inclusion in the trial, and on this visit a worsening of the clinical situation will be evaluated by the GP to determine if a change of therapy is necessary or not.

5. The discussion part would benefit from a detailed account of the strengths and weaknesses of the methodology, data collection instruments and analysis. It would be good as well if you highlight the expected impact of confirming the hypothesis of the study.
Response: We have included the strengths and limitations of the study separately from the text as suggested by the journal's guidelines. We have completed the discussion of the paper as suggested.

MINOR REVISION:
1. Page 4. Following the title, please write first the strengths and afterwards the limitations. Check that everything you wrote in this page is part of the discussion section
Response: We have changed the order, thank you.

2. Line 41 page 8 "an effect on their beliefs is expected". The beliefs of the doctors or the patients…. Or actually both?
Response: In both.

3. Page 5 line 55. Grammar issue: who is the subject of the action? The GP? Or the patient?
Response: Sorry for the misunderstanding. This sentence is now clearer.

4. I would suggest language revision by a native English speaker
Response: Thank you for the recommendation. A native English speaker has reviewed the whole manuscript.

Reviewer: 2
Reviewer Name: Lisa Bebell
Institution and Country: Massachusetts General Hospital and Harvard Medical School, USA
Please state any competing interests or state 'None declared': None declared
This protocol is designed to determine the safety of antibiotic discontinuation among outpatients originally prescribed antibiotics, but suspected to have viral diseases. The main hypothesis put forth by the authors is that although antibiotic discontinuation is recommended for suspected viral infections, there is no evidence that stopping antibiotics is safe. The authors argue that the Spanish Society of Family Medicine has been recommending GPs to ask their patients to stop taking antibiotics when they suspect a viral infection since 2011, but this recommendation is not evidence-based. Thus, they have designed a multicenter clinical trial protocol to determine the safety of this recommendation in adult patients.
The protocol is generally well-written. The subject is of broad interest to clinicians practicing in nearly

all settings, as this clinical situation is commonly encountered regardless of geographic location and resource availability. The main hypothesis is reasonable, though the extensive and vague exclusion criteria and high proportion of projected loss to follow-up concern me that the results will not be generalizable in a meaningful way.
I believe the paper would be substantially improved by increasing the specificity of the inclusion and exclusion criteria, decreasing the number of clinicians and centers involved, and providing the tools to be used for qualitative data collection. Major and minor points are described in detail below.
Major Points:
1. Patients must consent to participate in this open-label, non-blinded study. I am concerned that the population agreeing to participate will not be representative of the Spanish outpatient population at large, and therefore that results of this trial may not be generalizable enough to be useful in clinical practice. Specifically, patients (and clinicians) agreeing to participate in this trial are likely to be more open to the idea of stopping antibiotics in the case of suspected viral illness. Trial participants may be less likely to have taken non-prescribed antibiotics, or may be untruthful about their antibiotic use in a way not representative of your outpatient population. This could also introduce information bias and lead to a misinterpretation of the results.
Response: This is a good point which we want to make clear. As already mentioned in the text our local society of family medicine recommends doctors to discontinue antibiotic treatment when they think the use of antibiotics is pointless. However, very few doctors accomplish this recommendation as many are afraid of complications, worse prognosis, longer duration of symptoms, etc. if they do it. We acknowledge that many doctors cannot be recruited for this randomised clinical trial as they would feel uncomfortable (see also next query's answer). We can only include clinicians who think that discontinuing antibiotic treatment is feasible in suspected viral infections, but if the results of the trial are in line with our hypothesis we will find it easier to convince doctors to follow suit. When it comes to non-prescribed antibiotics we do not think that patients purchasing an antibiotic or taking a leftover antibiotic found in their households would be less represented in this trial than the overall outpatient population since this group of patients is large in our country.

2. How will the authors account for variability in experience, confidence and comfort levels of GPs in their assessment that "the GP considers that antibiotics are not needed to be taken"? It seems possible that some GPs in your study could have abnormally low or high confidence in their ability to assess whether antibiotics are necessary, this could skew the results significantly, and potentially lead to a greater number of adverse events. Perhaps inviting GPs who have a midrange level of experience (e.g. have been in practice for 5-15 years) might help to equalize their confidence levels? Another strategy would be to administer a questionnaire to the GPs using vignettes or questions to assess their level of confidence in stopping antibiotics, which might make your results more interpretable.
Response: As mentioned in the previous query, we intend to recruit only experienced GPs who feel comfortable in stopping antibiotic treatments in this trial, since our goal is the hypothesis that there are no differences in the number of days with severe and very severe symptoms between the two groups. We would like to convince other doctors who are still uncomfortable with this strategy to start doing this. We appreciate the suggestion of administering a questionnaire to the GPs with the utilisation of questions, guidelines and clinical vignettes to assess their level of confidence with this strategy and we have added this to the methodology of the study.

3. The authors propose secondary study objectives to study the incidence adverse events (page 9, objective 1) and number of complications within the first 3 months (page 9, objective 4). Though these are not primary objectives, I believe the study will be underpowered to detect most adverse events and complications, and this limitation should be recognized more clearly.
Response: We fully agree with this suggestion. The calculation of the sample size of the trial is based on the main outcome, and the number of complications and adverse effects are supposed to very low in this trial. We have added this limitation to the corresponding part, as follows: 'This trial might be

underpowered for the detection of differences between the two groups in terms of adverse events and complications within the first three months, since these outcomes are considered secondary endpoints in this study'.

4. The authors propose additional qualitative secondary study objectives to assess antibiotic consumption (page 9, objective 2) and assess satisfaction (page 9, objective 3). However, the tools used for these assessments are not provided, and the authors do not adequately acknowledge the potential for information bias when relying on patient self-report of these measures.
Response: We have decided to delete this part as our goal is to carry out a randomised clinical trial. Provided that discontinuing antibiotic treatment is safe, then we plan to set up a qualitative study once the trial has finished.

5. I am concerned that the authors predict a 20% proportion of enrollees will be lost to follow-up (page 10, lines 53-55). This is a high proportion lost, and these patients are likely to be significantly different to those retained in the study, including their compliance with antibiotic recommendations. Loss of this high proportion of participants could significantly affect trial results.
Response: The primary outcome of this study is provided in the symptom diaries. In previous European-based studies we observed a number of approximately 20% of diaries not returned. However, this percentage could be lowered with the use of shorter diaries and by encouraging the patients to reconsult once the implementation of the diaries has finished (visit number 3) as already suggested in the protocol. Thus, we have lowered this percentage to 15% as in other studies.

6. The authors plan to invite 40 clinicians from 20 different healthcare settings to join their study, anticipating that each clinician will recruit 10-12 participants over the 2-year study period to achieve their projected enrollment of 430 total participants. Though it is a noble goal to include clinicians and patients from diverse practice settings, I am concerned that it will not be feasible or practical to ensure that 40 clinicians can be trained and achieve compliance with the study protocol and recommendations, and retain these skills over a 2-year period while enrolling participants infrequently. Moreover, recruitment of 5-6 participants annually (1 every 2 months) represents a very small proportion of all patients seen in each provider's clinical practice. If, as the authors state, 15% of all patient visits are for viral conditions targeted by this trial, I would estimate that most clinicians would see approximately 1-3 patients every day with viral illnesses that might qualify them for this study. The expectation that each clinician would recruit such a small proportion of the eligible patient population concerns me that clinicians would 'cherry-pick' patients they wished to enroll, choosing to approach potential research subjects not at random, but rather in a way that is convenient to their clinical practice. In my opinion, this has great potential to introduce systematic bias into the sample, and invalidate the results. I think the trial would be more valid if it were restricted to a smaller number of clinicians (~10), and these clinicians were mandated to approach all eligible patients sequentially for enrollment, until they reached their target enrollment of 43 participants each. In addition, I would request the authors specify what data can be collected from participants who chose not to enroll, important information to help future readers decide how applicable trial results are to their own population.
Response: We have slightly increased the number of patients to be recruited in this trial (240 per group), and the patients to be included should also fulfil the requirement that the prior number of days taking antibiotics should be less than 3. As suggested by reviewer 3, there is increasing evidence on the effectiveness of short antibiotic therapy regimens (from 3 to 5 days with antibiotics) in some RTIs, with similar outcomes compared to longer therapies. Provided that the main results – days with severe symptoms - are similar in the two groups, our goal is to convince clinicians to stop antibiotic regimens if they no longer feel this is necessary.
As suggested in your comment, we have reconsidered the number of centres and clinicians in each centre, recruiting fewer GPs (15 in total), and they will be asked to consecutively recruit the first 32 participants with RTIs. As mentioned before in a previous comment, all the potential participants will

do a test with clinical cases and vignettes to better identify the best candidates for this trial. There will be a coordinator in each of the 10 participating centres, and they will be responsible for the good progress of the trial in the centres and for maintaining these skills.

Minor Points, page 5:
1.Page 5. Lines 31-32: I think it is an overstatement to state “General practitioners (GP) have always been told to continue an antibiotic regimen once the patient has initiated it in order to prevent the patient from acquiring resistant organisms.” Though I am unfamiliar with general practice in Spain, using the word always might be too strong. I recommend softening this statement to eliminate that word and instead state “Since 2011, general practitioners (GPs) in Spain have been told to continue . . .”
Response: We have softened the sentence as suggested. This often-heard dogma about completing an antibiotic course once initiated first appeared during our years studying medicine, and it is an overarching statement here in Spain. Notwithstanding, we have toned down this sentence by changing the word ‘always’ to ‘generally’.

2. Page 5. Lines 34-52: This section discusses the appropriate dosing of antibiotics for treatment of diagnosed bacterial infections. While the authors make valid points, I think this discussion is tangential to the focus of the protocol and could be eliminated or significantly shortened so as not to distract from the main focus of the paper, which is about antibiotic use in non-bacterial infections.
Response: We have reworded this paragraph as also suggested by another reviewer.

3. Page 5. Lines 54-56: Perhaps you are describing a local Spanish practice of making home visits to patients, but if you intend to describe office visits with patients, I recommend re-wording this sentence to say “GPs often see patients with suspected viral infections of the upper and lower airways, for which antibiotic treatment makes no difference in terms of clinical outcomes.” Furthermore, to state that antibiotic treatment makes no difference is a strong statement. Some antibiotics have known anti-inflammatory properties and may improve the outcome of viral illness by this mechanism. In addition, to my knowledge there is not enough evidence to support the statement, and in fact, a negative impact (e.g. an antibiotic-related side effect or new case of C. difficile diarrhea) could be viewed as a difference in clinical outcome. Please consider rewording this statement with this view in mind.
Response: We have changed the first sentence as suggested and toned down the second sentence.

Minor Points, page 6:
1. Page 6. Lines 4-5: Consider exchanging ‘bugs’ for ‘bacteria’.
Response: Changed.

2. Page 6. Lines 6-7: Why should this be common practice? I follow your argument, but there are many reasons why it should not be common practice, including many behavioral aspects of clinical medicine including the desire to give some type of treatment to the patient, patient demand of medications, etc. I recommend that you consider rewording this to take into account opposing views.
Response: We have changed this paragraph.

3. Page 6. Lines 11-13: This statement about GPs feeling unsafe to discontinue an antibiotic seems reasonable, but you do not support it with a citation. Can you provide a reference for this assertion?
Response: Yes, we have added two references which back up this statement.

4. Page 6. Lines 32-34: See my comments above about ‘no difference.’ Perhaps this could be reworded to state that antibiotics ‘do not improve outcomes’ for viral upper respiratory infections?
Response: We have changed this sentence accordingly.

5. Page 6. Lines 25-27: A proportion of 2.2% of all episodes of common cold receiving antibiotic

treatment does not seem high to me, and is what I would expect under the very best clinical practice. Is this a typographical error?
Response: We have deleted this sentence. What is remarkable is that more than 60% of the antibiotics are prescribed for episodes of acute bronchitis and acute rhinosinusitis.

6. Page 6. Lines 49-56: This is an overly long heading and slightly confusing. Can it be shortened and reworded to state the problem more clearly?
Response: We have deleted this heading, and it has been unified with the previous one.

Minor Points, page 7:
1.Page 7. Lines 3-22: The authors state that antibiotics “might” be required to treat these infections, but that “most” are self-limiting. In my mind, this is the kind of statement that drives antibiotic use, because clinicians are not certain whether they are in the “might” category or the “most” category. I note this as a potential barrier to implementing your trial, and also because it seems to me a rational (rather than irrational) reason for antibiotic use that may be hard to address. It might be helpful to reword this section to address this conundrum, and set up the reader to understand how you will address this particular problem in your trial.
Response: The whole paragraph has been reworded.

Minor Points, page 9:
1.Page 9. Line 18: Please change ‘and’ to ‘or’ for clarity.
Response: Done.

2. Page 9. Lines 32-34: How will you assess for an ‘inadequate family setting’? Do you mean homelessness? Or something else? These criteria seem vague and might benefit from more specificity.
Response: You are right; we have deleted this part of the sentence.

3. Page 9. Lines 45-46: It might help to specify the timeline for a ‘terminal disease’, e.g. life expectancy less than 6 or 12 months.
Response: Done.

4. Page 9. Lines 49-50: Similarly to above, it is vague what you mean by ‘difficulty to attend the programmed visits’. Can you specify a distance from clinic or another measure that would clarify this exclusion?
Response: We have changed this sentence. Patients who state that they are unable to see their doctor at the practice will be excluded.

Minor Points, page 12:
1. Page 12. Lines 19-20: The authors state they will use Fisher’s exact test to compare outcomes in the randomized groups. This test may be difficult to use given the number of proposed participants (430). A Chi2 test may be more appropriate, when estimated cell sizes are >5. 2.
Response: Changed.

2. Page 12. Lines 36-39: One of the criteria for withdrawal is “unsatisfactory therapeutic effect”. This is vague and could lead to protocol discontinuation for a number of reasons. Can you be more specific, so the reader can understand how this will not introduce bias? I am concerned that participants in the intervention group will be dissatisfied with ongoing symptoms after being told to stop antibiotics, and will withdraw from the study in greater numbers than those in the control group, affecting your results.
Response: This term has been deleted from the protocol. Since the duration of symptoms and ‘severe symptoms’ will be based on what the patients write in the symptom diaries, and GPs will decide 2-3 days after the basal visit if there is a worsening of the clinical situation, we preferred to delete this

part.

Reviewer: 3
Reviewer Name: Gwendolyn (Lyn) Gilbert
Institution and Country: University of Sydney, Australia
Please state any competing interests or state ‘None declared’: None declared
Most of my "yes" and "no" answers above should be qualified by "partially", "maybe" etc. I hope the following comments clarify them.
The paper describes an open label randomised controlled trial of the safety of discontinuing patient antibiotic treatment when the physician does not consider it necessary. It aims to test the common fear among GPs that it is harmful to stop a course of antibiotics once started, even if it was not needed in the first place. Whilst this may be a common belief, it has never been clear why it should be harmful in these circumstances. Presumably the fear is that the judgement, that an antibiotic was not needed, was incorrect. However, the same would apply to the decision not to start antibiotics.
The only plausible adverse outcome of stopping an antibiotic prematurely is that symptoms would persist or recur if the decision was wrong – in which case it can be started or restarted for the small proportion of patients affected. There is no evidence that stopping the antibiotic before the course is finished will increase the risk of acquiring a resistant organism – on the contrary, there is evidence that the longer the course the greater the risk. Moreover, the length of a course of antibiotic is generally quite arbitrary and the relatively few studies comparing shorter, with “conventional” courses usually demonstrate equivalence.
On the other hand there are many plausible benefits of stopping or not starting unnecessary antibiotic therapy for the majority of patients with viral infections who did not need it. Based on systematic reviews cited by the authors the proportion of patients, in whom symptoms persist for longer when antibiotics were not given for sore throat or acute rhinosinusitis, is significant but very small and the additional duration of symptoms short. Presumably these differences would be even less in the group of patients to be studied in this trial, who have already started antibiotics, depending on how long they have been taking them. Two or three days of antibiotics might be as adequate a “course” as a full 5 or 7 day conventional course.
These considerations are not well presented in the abstract or background and the manuscript would benefit from some elaboration of these points.
Response: Thank you very much for these remarks. We have significantly changed the wording of the Introduction section taking your comments into account. We have added information about the ‘similar efficacy’ of short antibiotic therapy courses (from 3 to 5 days) in some RTIs compared to longer duration therapies and the lack of studies comparing discontinuation vs. continuation of therapies. We have slightly changed the inclusion criteria and in order to differentiate this trial from the short vs. long duration therapies in some of these RTIs, we now plan to recruit only patients who had been taking antibiotics for less than 3 days. Our goal is to convince GPs that discontinuing antibiotic therapy is not harmful, provided the results are equivalent in the two groups.

On page 6 the authors state that there is no evidence that it is safe to interrupt a course of antibiotics that the GP has judged to be unnecessary. On the other hand there is also no evidence that this practice is, and many plausible reasons why is would not be, unsafe.
Response: We think that this point is now clearer in the new version of the manuscript.

On page 7, it would be useful to elaborate (with some data from the abstracts of the cited papers) on the statement “recent systematic reviews have suggested……etc.” (lines 3-4).
Response: Done.

Page 8. Study design. I wonder whether it is right to describe the control group as the one being treated by the “usual strategy”, when the practice of continuing the antibiotic when the GP thinks it is no longer necessary is specifically discouraged by the Spanish Society of Family Medicine? I

understand that this is contrary to the other (unfounded) belief that it is dangerous to stop a course, once started and that GPs are understandably confused, but wonder what proportion of “GPs (actually) are reluctant and fell unsafe to discontinue……..” if the GP her/himself believes it is unnecessary and especially if s/he did not prescribe it in the first place.
Response: The reviewer is right. Despite this recommendation, the strategy of discontinuing antibiotic therapy when the clinician no longer considers this not necessary is seldom used in our country. What clinicians usually do is allow the patient to continue this therapy even if they think they do not need it. Therefore, we have decided to eliminate the terms ‘control group’ and ‘intervention group’ in this resubmission, as they make no sense in this trial.

Page 9, lines 3-4. The eligibility criterion “……..patients feel that the antibiotic regimen has not worked as expected and fell they need clinical reassessment” is not clear. If they are randomised to the control group will they be continued on the antibiotic they have already started or changed to another one (and of the latter, will the same antibiotic regimen reused for all patients in the control group in all practices)?
Response: We have unified the first two clinical scenarios. We think that the description of the two sources of patients in this trial is clearer now.

For patients in the control group, how will the “course” of antibiotic to be completed be defined? What if they are taking a leftover antibiotic that is inappropriate or out of date,? Will they finish what they have in the cupboard at home or start a fresh course? If the latter, which antibiotic will be prescribed and will they take only as much of the new course to make up the equivalent of a conventional course? Or will they start and complete a new course of the same or a different antibiotic – which will mean a longer than usual total course?
Response: Thank you for this remark as it is very important. Previous use of antibiotics by patients before they come to see us is very common here. Patients who made the decision to take antibiotics by themselves (either purchased at the pharmacy or taken from leftovers stored at home) assigned to the usual strategy of continuing antibiotic treatment will be provided with a medical prescription of the same antibiotic until completing the recommended therapy duration according to local guidelines (7 days at least), even if the antibiotic is not first-line treatment.

If patients in the intervention group have continued symptoms on review – what will be the criteria for recommencing antibiotic therapy and how will they be standardised – number of symptoms, severity? With which antibiotic - the one they had already started initially or another chosen by the GP? Will it be the same for all intervention group patients?
Response: This is also a good point. It will be up to the physician to decide on the use of an antibiotic because of failure with a previous one, but this can occur in either of the two groups. The antibiotic used in case of failure will be the first-line drug recommended by the local guidelines. As mentioned in the new text, ‘patients will be interviewed by telephone 2 or 3 days after their inclusion in the study. At this first follow-up visit, a worsening of the clinical situation of the patient will be evaluated to determine whether antibiotic treatment is necessary among patients in the group assigned to discontinuation (first-line antibiotics will be recommended in this case) or whether the antibiotic regimen should be continued or also changed to the first-line drug in patients in the group allocated to continuation.’

Data analysis: page 12. I am not a statistician, but such a relatively small group of subjects and such a large number of different conditions, symptoms and severity gradings might make meaningful analysis difficult. Will severity gradings be equivalent across all different symptoms and conditions?
Response: Yes. Only symptoms with scores of 5 or 6 will be considered as severe and they will be the same for all the symptoms in this trial. We have eliminated the conditions of acute exacerbations of mild-to-moderate COPD as we were not expecting to have a large number of these patients in the trial and in order to lower the number of conditions.

What is the definition of a serious adverse event or unsatisfactory therapeutic effect? Do serious adverse events include complications of untreated infections and adverse effects of antibiotic therapy?
Response: Thank you for this comment. A serious adverse effect is defined as any untoward medical occurrence that results in death, is life-threatening, requires inpatient hospitalisation or prolongs existing hospitalisation, results in persistent or significant disability/incapacity, or requires intervention to prevent permanent impairment or damage. We have deleted the unsatisfactory therapeutic effect in this paragraph.

If patients in control group are able to interrupt medication during the study can those in the intervention group recommence the antibiotic they had started themselves before entry into the study?
Response: You are right. Patients will be allocated to either of the two strategies: discontinuing or continuing antibiotic treatment. However, both groups will be asked to write in the symptom diary if they had actually taken antibiotics or not.

Finally, there are a number of minor editorial/grammatical issues:
P 6 line 1. “bugs” replace with “bacteria”
Response: Changed.

p11, lines 21. “change or commence antibiotic treatment”
line 28. “…need to commence, change, continue or cease the antibiotic treatment ……..” Presumably a few patients will develop antibiotic side-effects that necessitate stopping prematurely.
Response: Changed.

P13, line 23 “…..withhold the discontinuation of….” Is awkward and confusing suggest “continue”
Response: Changed.

p14, lines 1-3. The sentence beginning “A post-trial implementation….” is confusing – needs rewording.
Response: We have deleted this part, as this does not correspond with the clinical trial.

Reviewer: 4
Reviewer Name: Jesse Berlin
Institution and Country: Johnson & Johnson, USA
Please state any competing interests or state ‘None declared’: I'm a full-time employee of Johnson & Johnson. I don't believe there's a direct conflict in reviewing the methodology being proposed in this protocol.
1. Throughout the protocol, there are places where the English (grammar and style) could use some attention.
Response: Thank you for your comment. A native English person has now reviewed the whole manuscript.

2. I’ll have mostly conceptual questions below. I won’t have a lot of specific detailed suggestions. (Just letting you know.)
Abstract: This is more of a general question. I would just ask whether “safety” is the right word to describe your primary objective. Does longer duration of severe symptoms (in terms of number of days) really meet the definition of a “safety” issue? I would think of it more as an efficacy (or reduced efficacy) problem. It’s a semantic point for your consideration, but when I think of safety, from a regulatory/regulated industry perspective, I tend to think more about actual adverse events. These

could result from inadequate treatment, so I don't feel strongly about making any changes. Again, for your consideration.
Response: This is an excellent point and the reviewer is absolutely right. We have slightly changed the title and we have included 'efficacy and safety'. In the previous manuscript, we had focused on the safety of discontinuing antibiotic therapy, but as the reviewer suggests, the duration of severe symptoms is more about efficacy rather than safety. The new title is: "The STOP-AB trial protocol: Efficacy and safety of discontinuing patient antibiotic treatment when physicians no longer consider it necessary."

3. The biggest challenge I had was understanding how you will operationalize the study, including how to define (in practice) the eligible population and how to get clinicians to enroll their patients. You say,
"We will include patients from 18 to 75 years of age with uncomplicated acute respiratory tract infections (RTIs) who have previously taken any dose of antibiotic but physicians no longer consider it necessary."
I'm not a clinician, so I may not appreciate or understand what happens in practice (other than having gone to my doctor with a sinus problem in the past). My apologies if I'm missing your point, but here is my problem. You're saying that the clinician will somehow recognize that a patient is inappropriately receiving an antibiotic. Basically the entry criterion is that the clinician says "this person should not have gotten an antibiotic." For the trial to be successful, the clinician needs to be in equipoise with respect to continuing the antibiotic. That seems to conflict directly with the premise that this same clinician recognizes the inappropriate nature of the original prescription (which he or she may have written). They wouldn't be considering discontinuation if they didn't believe it would be inappropriate to continue. Will clinicians really be willing to randomize patients? Would they say, "this patient is inappropriately on an antibiotic but I'm afraid to stop in any case?"
Response: I think you might have misunderstood this point because it may not have been clearly described before. We have reworded this part now. As mentioned in the methodology there will be two sources of patients: on one hand, patients who have initiated an antibiotic course themselves, and on the other hand, patients who had been prescribed an antibiotic by another health professional and return to the GP who no longer considers this treatment to be necessary. All the patients will be randomised with the use of an online randomisation platform thereby guaranteeing this equipoise.

4. Again, maybe I'm misunderstanding. Having said what I just did, it's possible that a clinician could decide to stop treatment with the antibiotic because of declining symptoms. In that case, I could see that clinician being in equipoise. If that's the situation you have in mind, maybe it would be helpful to stratify the randomization by reason for stopping (symptoms improved so maybe I can stop vs. condition not improving so maybe it's a viral infection vs. the culture results came back showing the infection is NOT bacterial?) Maybe there are different strata than these, but I hope you get the idea. I think things could look very different depending on why the clinician thinks it MIGHT be OK to stop.
Response: No, you didn´t misunderstand this point and this comment is very useful. We have slightly changed the inclusion criteria and in order to prevent clinicians from including patients with declining symptoms and considering the recent evidence that short therapy courses (from 3 to 5 days of antibiotic regimens) could be as effective as longer therapies, we are planning to only recruit patients who had taken antibiotics for less than three days. If the outcomes between the two groups are equivalent the evidence for discontinuing antibiotic therapies when they are no longer necessary will be more compelling.

5. You use duration of symptoms as the outcome measure. Duration from what time point? What is time zero? Is it the time of randomization? Time at which the antibiotic is actually stopped makes sense in the "stop treating" arm but there's no equivalent (except the time of randomization, I think) in the control (continue treatment) arm. This is a matter of minor clarification, I think.

Response: You are right, since the duration of symptoms can only be monitored from the day the patient comes to the consultation and is randomised. Obviously, we will collect information about the prior duration of symptoms in the basal visit.

6. Did you pilot test the symptom diaries? Will subjects be compliant? This diary is the major outcome measure.
Response: Yes. We used these symptom diaries in a previous clinical trial about delayed prescription of antibiotics (reference number 37). The diaries are short (as mentioned in Table 1) and take less than two minutes to complete every day. We observed a low percentage of diaries lost in this trial (less than 10%) and therefore, in the present clinical trial we will use the same strategies as those used in the previous trial.

7. Secondary objective: You list assessing the incidence of adverse effects of medication as a secondary objective. Technically, that's only possible in the continued treatment arm, by definition, no? Adverse EVENTS (not necessarily EFFECTS of the drug) could occur in either arm. (The exception, I suppose, would be delayed drug effects.) In any case, I think some additional detail would be helpful here.
Response: Patients in both groups will be treated by the GPs with the most appropriate therapy, except for antibiotic therapy. We do not expect that the use of other non-antibiotic therapies, such as expectorants, mucolytics, NSAIDs, paracetamol, β2-agonist inhalers, etc. would be different in the two groups, but we will consider all the adverse effects of the different therapies used by GPs in the whole sample of participants in this clinical trial.

8. ON A RELATED POINT: I wasn't quite clear about your exclusion criterion: "Lack of tolerance to oral treatment, such as the presence of nausea and vomiting, gastrectomy, post-surgery and/or diarrhea." Are you saying that if someone is having the antibiotic stopped specifically because of an adverse event, i.e., the patient PRESENTS with what is "clearly" (highly likely to be) an adverse effect of the drug, that should disqualify that person from participating in the trial because it would be unethical to continue treatment in that person. That's fine, and I agree with the rationale (if I'm getting it correct), but again, some clarification will be helpful. As a relevant aside, if you are excluding people who had an adverse EFFECT (here I do mean effect) of the drug, that will reduce the probability that people who are randomized to continued treatment will have a subsequent adverse effect from the drug.
Response: Yes, we agree. Patients who had had an adverse effect due to the antibiotic initially prescribed cannot be recruited in this trial. It wouldn't be fair to continue with the same antibiotic which had caused this adverse effect. However, by limiting the days of previous antibiotic therapy a patient can be recruited - less than 3 days -, the number of previous adverse effects should be low.

9. Can you define "complications related to the RTI?" How do you distinguish these from other adverse events or symptoms?
Response: We will consider the cases of pneumonia, empyema, peritonsillar abscess, mastoiditis, otitis media, bacterial meningitis, and intracranial abscess as in the recent study of Gulliford et al published in the BMJ.

10. I'm assuming the sample size is TOTAL? Not per group? Based on the SD of 3.3? It's a picky technical point, but there's an underlying normality assumption if you're basing your calculation on the t-test (which is what the statistical analysis section indicates). One could argue for some other (e.g., Poisson distribution) approach, as you are counting days, and the Poisson distribution is often used to model counts. It seems very likely that the distribution of duration will be asymmetric (highly skewed). If the mean is 3.6 and SD is 3.3, there is likely a long tail to the right. Similarly – is the t-test the right test for comparing duration?
Response: The main outcome is the duration of 'severe' and 'very severe' symptoms. Therefore, we

do not expect a highly asymmetric distribution of this main variable, as we didn't observe this in the trial these figures have been taken from. However, the reviewer is right that this test might not be particularly appropriate in the case of a highly-skewed distribution of this variable, but we do not expect this.

We hope the protocol has suitably improved to be accepted for publication.

VERSION 2 – REVIEW

REVIEWER	Gloria Cordoba Department of General Practice University of Copenhagen Denmark I know personally two of the authors
REVIEW RETURNED	17-Mar-2017

GENERAL COMMENTS	The authors have addressed all the points highlighted by all the reviewers. The research question is very relevant and the pragmatic design with its advantages and disadvantages is currently the best approach to answer the research question in the general practice setting.

REVIEWER	Lisa Bebell Massachusetts General Hospital, USA
REVIEW RETURNED	22-Mar-2017

GENERAL COMMENTS	Please see attachment for a better-formatted version of these comments. This protocol is designed to determine the safety of antibiotic discontinuation among outpatients originally prescribed antibiotics, but suspected to have viral diseases. The main hypothesis put forth by the authors is that although antibiotic discontinuation is recommended for suspected viral infections, there is no evidence that stopping antibiotics is safe. The authors argue that the Spanish Society of Family Medicine has been recommending GPs to ask their patients to stop taking antibiotics when they suspect a viral infection since 2011, but this recommendation is not evidence-based. Thus, they have designed a multicenter clinical trial protocol to determine the safety of this recommendation in adult patients. The original protocol was returned to the investigators for revision, and I am now reviewing the revised version. The protocol is generally well-written. The subject is of broad interest to clinicians practicing in nearly all settings, as this clinical situation is commonly encountered regardless of geographic location and resource availability. The main hypothesis is reasonable, though the extensive and vague exclusion criteria and high proportion of projected loss to follow-up concern me that the results will not be generalizable in a meaningful way. I believed the original protocol would be substantially improved by increasing the specificity of the inclusion and exclusion criteria, decreasing the number of clinicians and centers involved, and

	providing the tools to be used for qualitative data collection. My original major and minor points are described in detail below, and following those points are my assessment of whether they have been addressed. Major Points: 1. Patients must consent to participate in this open-label, non-blinded study. I am concerned that the population agreeing to participate will not be representative of the Spanish outpatient population at large, and therefore that results of this trial may not be generalizable enough to be useful in clinical practice. Specifically, patients (and clinicians) agreeing to participate in this trial are likely to be more open to the idea of stopping antibiotics in the case of suspected viral illness. Trial participants may be less likely to have taken non-prescribed antibiotics, or may be untruthful about their antibiotic use in a way not representative of your outpatient population. This could also introduce information bias and lead to a misinterpretation of the results. Informed consent is retained in the current proposal, however, these limitations are acknowledged by the investigators, especially the open-label design which may favor placebo in the antibiotic-receipt group. 2. How will the authors account for variability in experience, confidence and comfort levels of GPs in their assessment that “the GP considers that antibiotics are not needed to be taken”? It seems possible that some GPs in your study could have abnormally low or high confidence in their ability to assess whether antibiotics are necessary, this could skew the results significantly, and potentially lead to a greater number of adverse events. Perhaps inviting GPs who have a midrange level of experience (e.g. have been in practice for 5-15 years) might help to equalize their confidence levels? Another strategy would be to administer a questionnaire to the GPs using vignettes or questions to assess their level of confidence in stopping antibiotics, which might make your results more interpretable. The authors have taken this concern into account and decreased the number of centers involved from 10 to 20. This will help somewhat to reduce heterogeneity, though heterogeneity of the study population and the clinicians may still affect the interpretability of the results. 3. The authors propose secondary study objectives to study the incidence adverse events (page 9, objective 1) and number of complications within the first 3 months (page 9, objective 4). Though these are not primary objectives, I believe the study will be underpowered to detect most adverse events and complications, and this limitation should be recognized more clearly. The likelihood that the study will be underpowered has been addressed by the authors as a weakness. The sample size of 240 per group is justified, but I am still concerned that other outcomes may not be assessable with this relatively small sample. 4. The authors propose additional qualitative secondary study objectives to assess antibiotic consumption (page 9, objective 2) and assess satisfaction (page 9, objective 3). However, the tools used for these assessments are not provided, and the authors do not adequately acknowledge the potential for information bias when

	relying on patient self-report of these measures. The questionnaire has now been included, and the authors note that it has been validated. The authors account for 15% loss of participants due to diary non-completion. A word of caution that those participants not completing the diary will likely differ from those who do complete the diary. I recommend analyzing the demographics and other data from these participants who do not complete the diary (whom I understand will not be included in the analysis) to determine if there are significant differences between these participants and others that could introduce bias. 5. I am concerned that the authors predict a 20% proportion of enrollees will be lost to follow-up (page 10, lines 53-55). This is a high proportion lost, and these patients are likely to be significantly different to those retained in the study, including their compliance with antibiotic recommendations. Loss of this high proportion of participants could significantly affect trial results. See my comments in #4, above. Though the investigators have decreased the proportion lost from 20% → 15% (for unclear reasons, though a prior study is now cited), I still worry that this could be a source of bias. 6. The authors plan to invite 40 clinicians from 20 different healthcare settings to join their study, anticipating that each clinician will recruit 10-12 participants over the 2-year study period to achieve their projected enrollment of 430 total participants. Though it is a noble goal to include clinicians and patients from diverse practice settings, I am concerned that it will not be feasible or practical to ensure that 40 clinicians can be trained and achieve compliance with the study protocol and recommendations, and retain these skills over a 2-year period while enrolling participants infrequently. Moreover, recruitment of 5-6 participants annually (1 every 2 months) represents a very small proportion of all patients seen in each provider's clinical practice. If, as the authors state, 15% of all patient visits are for viral conditions targeted by this trial, I would estimate that most clinicians would see approximately 1-3 patients every day with viral illnesses that might qualify them for this study. The expectation that each clinician would recruit such a small proportion of the eligible patient population concerns me that clinicians would 'cherry-pick' patients they wished to enroll, choosing to approach potential research subjects not at random, but rather in a way that is convenient to their clinical practice. In my opinion, this has great potential to introduce systematic bias into the sample, and invalidate the results. I think the trial would be more valid if it were restricted to a smaller number of clinicians (~10), and these clinicians were mandated to approach all eligible patients sequentially for enrollment, until they reached their target enrollment of 43 participants each. In addition, I would request the authors specify what data can be collected from participants who chose not to enroll, important information to help future readers decide how applicable trial results are to their own population. This concern has been partially addressed. There will be a fewer number of centers included with a larger number of participants at each center. However, the number enrolled still represent a small proportion of those seen at each center. I would advocate for decreasing the number of centers involved further, in order to have more participants seen by fewer clinicians. However, this is clearly a

	tradeoff between consistency and generalizability, and the investigators are in the best position to determine what will be most informative. Minor Points: Page 5 1. Lines 31-32: I think it is an overstatement to state “General practitioners (GP) have always been told to continue an antibiotic regimen once the patient has initiated it in order to prevent the patient from acquiring resistant organisms.” Though I am unfamiliar with general practice in Spain, using the word always might be too strong. I recommend softening this statement to eliminate that word and instead state “Since 2011, general practitioners (GPs) in Spain have been told to continue . . .” This language has been softened. 2. Lines 34-52: This section discusses the appropriate dosing of antibiotics for treatment of diagnosed bacterial infections. While the authors make valid points, I think this discussion is tangential to the focus of the protocol and could be eliminated or significantly shortened so as not to distract from the main focus of the paper, which is about antibiotic use in non-bacterial infections. This has been addressed. 3. Lines 54-56: Perhaps you are describing a local Spanish practice of making home visits to patients, but if you intend to describe office visits with patients, I recommend re-wording this sentence to say “GPs often see patients with suspected viral infections of the upper and lower airways, for which antibiotic treatment makes no difference in terms of clinical outcomes.” Furthermore, to state that antibiotic treatment makes no difference is a strong statement. Some antibiotics have known anti-inflammatory properties and may improve the outcome of viral illness by this mechanism. In addition, to my knowledge there is not enough evidence to support the statement, and in fact, a negative impact (e.g. an antibiotic-related side effect or new case of C. difficile diarrhea) could be viewed as a difference in clinical outcome. Please consider rewording this statement with this view in mind. This has been revised. Page 6 1. Lines 4-5: Consider exchanging ‘bugs’ for ‘bacteria’. Addressed. 2. Lines 6-7: Why should this be common practice? I follow your argument, but there are many reasons why it should not be common practice, including many behavioral aspects of clinical medicine including the desire to give some type of treatment to the patient, patient demand of medications, etc. I recommend that you consider rewording this to take into account opposing views. Addressed. 3. Lines 11-13: This statement about GPs feeling unsafe to discontinue an antibiotic seems reasonable, but you do not support it with a citation. Can you provide a reference for this assertion?

	Addressed. 4. Lines 32-34: See my comments above about 'no difference.' Perhaps this could be reworded to state that antibiotics 'do not improve outcomes' for viral upper respiratory infections? Addressed. 5. Lines 25-27: A proportion of 2.2% of all episodes of common cold receiving antibiotic treatment does not seem high to me, and is what I would expect under the very best clinical practice. Is this a typographical error? This has been removed. 6. Lines 49-56: This is an overly long heading and slightly confusing. Can it be shortened and reworded to state the problem more clearly? Addressed. Page 7 1. Lines 3-22: The authors state that antibiotics "might" be required to treat these infections, but that "most" are self-limiting. In my mind, this is the kind of statement that drives antibiotic use, because clinicians are not certain whether they are in the "might" category or the "most" category. I note this as a potential barrier to implementing your trial, and also because it seems to me a rational (rather than irrational) reason for antibiotic use that may be hard to address. It might be helpful to reword this section to address this conundrum, and set up the reader to understand how you will address this particular problem in your trial. This section has been revised. Page 9 1. Line 18: Please change 'and' to 'or' for clarity. Addressed. 2. Lines 32-34: How will you assess for an 'inadequate family setting'? Do you mean homelessness? Or something else? These criteria seem vague and might benefit from more specificity. This has been removed/revised. 3. Lines 45-46: It might help to specify the timeline for a 'terminal disease', e.g. life expectancy less than 6 or 12 months. Addressed. 4. Lines 49-50: Similarly to above, it is vague what you mean by 'difficulty to attend the programmed visits'. Can you specify a distance from clinic or another measure that would clarify this exclusion? Addressed. Page 12

	1. Lines 19-20: The authors state they will use Fisher's exact test to compare outcomes in the randomized groups. This test may be difficult to use given the number of proposed participants (430). A Chi2 test may be more appropriate, when estimated cell sizes are >5. Addressed. 2. Lines 36-39: One of the criteria for withdrawal is "unsatisfactory therapeutic effect". This is vague and could lead to protocol discontinuation for a number of reasons. Can you be more specific, so the reader can understand how this will not introduce bias? I am concerned that participants in the intervention group will be dissatisfied with ongoing symptoms after being told to stop antibiotics, and will withdraw from the study in greater numbers than those in the control group, affecting your results. Largely addressed, analysis will be conducted using an ITT model. New minor points: 1. Page 1, abstract introduction, lines 9-10: I think the rationale that GPs don't stop antibiotics because they are not sure doing so is safe is a weak one. I think the stronger reasons for failure to stop antibiotics in the face of a viral illness are pressures from patients to continue giving some form of therapy and lack of confidence that an illness is truly viral, without any bacterial component. I think this is worth rewording in the abstract and the protocol to clearly state the rationale. 2. Background, page 5, lines 41-49: There is new information here, including a discussion of sort-course antibiotics. It would be helpful to cite evidence that GPs are not using short courses for pneumonia (which are becoming more common in the US). Also, the analogy between UTI and other types of infection isn't very helpful, since the pathophysiology of each infection and each anatomic site is markedly different. 3. Please revise lines 18-20 on page 6 "nobody wants to be seen" to make this a less casual and more formal statement. This manuscript would benefit from a review by a native English speaker with expertise in scientific writing. 4. Page 6, lines 27-34: please see my comment #1 above. I think it is important to note that the physician-patient dynamic also places pressure on doctors to continue antibiotics, in order to be seen as 'doing something' for their patient. 5. Page 11, line 45: the 'basal' visit should be changed to 'baseline' visit, I think.

REVIEWER	Gwendolyn (Lyn) Gilbert University of Sydney Australia
REVIEW RETURNED	20-Mar-2017

GENERAL COMMENTS	The protocol os much improved and previous ambiguities clarified.

REVIEWER	Jesse Beriln Johnson & Johnson, Global Epidemiology

	I'm a full-time employee of Johnson & Johnson, which does market antibiotics, but I don't see any conflict with a study that proposes to examine a facet of appropriate use.
REVIEW RETURNED	16-Mar-2017

GENERAL COMMENTS	I thank the authors for a thorough and thoughtful review to earlier comments. I have one additional comment and one statement of agreement with an important decision you made. 1. In the protocol, you now say you will collect “Adverse events of the medication given for this infection reported by the patients.” In your response to my earlier comments related to this, you said: “We do not expect that the use of other non-antibiotic therapies, such as expectorants, mucolytics, NSAIDs, paracetamol, β2-agonist inhalers, etc. would be different in the two groups, but we will consider all the adverse effects of the different therapies used by GPs in the whole sample of participants in this clinical trial.” I realize now that my earlier comment was not quite on target, for which I apologize. Your two statements (from the protocol and from your response) seem to conflict with each other. Here’s my thinking on this. Some patients will experience AEs, which could be caused by the antibiotic (particularly in the “continue treatment” group), or by another medication, or by the underlying illness or a comorbid condition. The randomization should tend to equalize the use of concomitant medications and comorbidities (at least at baseline). All of this makes attribution unreliable (i.e., physicians may not be able to tell whether a particular AE is related to the antibiotic or another factor). I would suggest that you collect all AEs and report them all in an appropriate format. If you want to collect attribution by physician (“related to antibiotic or not”) that’s fine, but my personal view, as I said, is that such attribution isn’t reliable. Again, my apologies for a confusing earlier comment. 2. You are limiting consideration to very experienced physicians from a modest number of clinical sites. I think, at this point in the understanding of antibiotic discontinuation, this is a good strategy. I think your aim is to make this similar in spirit to an “efficacy” study of a drug, i.e., under well-controlled conditions, we want to know if the intervention works. (I’m agreeing with this change).

VERSION 2 – AUTHOR RESPONSE

Reviewer: 4
Reviewer Name: Jesse Beriln
Institution and Country: Johnson & Johnson, Global Epidemiology
I thank the authors for a thorough and thoughtful review to earlier comments. I have one additional comment and one statement of agreement with an important decision you made.
1. In the protocol, you now say you will collect “Adverse events of the medication given for this infection reported by the patients.” In your response to my earlier comments related to this, you said: “We do not expect that the use of other non-antibiotic therapies, such as expectorants, mucolytics, NSAIDs, paracetamol, β2-agonist inhalers, etc. would be different in the two groups, but we will consider all the adverse effects of the different therapies used by GPs in the whole sample of participants in this clinical trial.” I realize now that my earlier comment was not quite on target, for

which I apologize. Your two statements (from the protocol and from your response) seem to conflict with each other.
Here's my thinking on this. Some patients will experience AEs, which could be caused by the antibiotic (particularly in the "continue treatment" group), or by another medication, or by the underlying illness or a comorbid condition. The randomization should tend to equalize the use of concomitant medications and comorbidities (at least at baseline). All of this makes attribution unreliable (i.e., physicians may not be able to tell whether a particular AE is related to the antibiotic or another factor). I would suggest that you collect all AEs and report them all in an appropriate format. If you want to collect attribution by physician ("related to antibiotic or not") that's fine, but my personal view, as I said, is that such attribution isn't reliable.
Again, my apologies for a confusing earlier comment.
Response: The reviewer is right. We have amended and simplified this sentence as follows: 'Adverse events in the two study arms'.

2. You are limiting consideration to very experienced physicians from a modest number of clinical sites. I think, at this point in the understanding of antibiotic discontinuation, this is a good strategy. I think your aim is to make this similar in spirit to an "efficacy" study of a drug, i.e., under well-controlled conditions, we want to know if the intervention works. (I'm agreeing with this change).
Response: Thank you for your comments.

Reviewer: 1
Reviewer Name: Gloria Cordoba
Institution and Country: Department of General Practice, University of Copenhagen, Denmark
The authors have addressed all the points highlighted by all the reviewers. The research question is very relevant and the pragmatic design with its advantages and disadvantages is currently the best approach to answer the research question in the general practice setting.
Response: Thank you.

Reviewer: 3
Reviewer Name: Gwendolyn (Lyn) Gilbert
Institution and Country: University of Sydney, Australia
The protocol is much improved and previous ambiguities clarified.
Response: Thank you.

Reviewer: 2
Reviewer Name: Lisa Bebell
Institution and Country: Massachusetts General Hospital, USA
Please state any competing interests or state 'None declared': None
Please see attachment for a better-formatted version of these comments. This protocol is designed to determine the safety of antibiotic discontinuation among outpatients originally prescribed antibiotics, but suspected to have viral diseases. The main hypothesis put forth by the authors is that although antibiotic discontinuation is recommended for suspected viral infections, there is no evidence that stopping antibiotics is safe. The authors argue that the Spanish Society of Family Medicine has been recommending GPs to ask their patients to stop taking antibiotics when they suspect a viral infection since 2011, but this recommendation is not evidence-based. Thus, they have designed a multicenter clinical trial protocol to determine the safety of this recommendation in adult patients. The original protocol was returned to the investigators for revision, and I am now reviewing the revised version. The protocol is generally well-written. The subject is of broad interest to clinicians practicing in nearly all settings, as this clinical situation is commonly encountered regardless of geographic location and resource availability. The main hypothesis is reasonable, though the extensive and vague exclusion criteria and high proportion of projected loss to follow-up concern me that the results will not be generalizable in a meaningful way.

I believed the original protocol would be substantially improved by increasing the specificity of the inclusion and exclusion criteria, decreasing the number of clinicians and centers involved, and providing the tools to be used for qualitative data collection. My original major and minor points are described in detail below, and following those points are my assessment of whether they have been addressed.

Major Points:

1. Patients must consent to participate in this open-label, non-blinded study. I am concerned that the population agreeing to participate will not be representative of the Spanish outpatient population at large, and therefore that results of this trial may not be generalizable enough to be useful in clinical practice. Specifically, patients (and clinicians) agreeing to participate in this trial are likely to be more open to the idea of stopping antibiotics in the case of suspected viral illness. Trial participants may be less likely to have taken non-prescribed antibiotics, or may be untruthful about their antibiotic use in a way not representative of your outpatient population. This could also introduce information bias and lead to a misinterpretation of the results.

Informed consent is retained in the current proposal, however, these limitations are acknowledged by the investigators, especially the open-label design which may favor placebo in the antibiotic-receipt group.

Response: Thank you for this remark. We acknowledge that the clinicians participating in this trial will not be representative of the community of GPs in our country, since only experienced doctors who are comfortable with the study will be invited to participate, but we honestly think that the patients to be recruited in this study will actually be representative of the patients attending the GPs' consultations. All the patients who meet the inclusion criteria and do not have any of the exclusion criteria will be asked to participate in the trial, and GPs will be obliged to fill out a screening log with all the patients who meet these criteria regardless of whether they consent or not: thus, if patients do not agree to participate we will be able to know why they decline to participate. We have added this information to the Methods section, as follows: 'GPs will fill out a screening log with all the patients who meet all the inclusion criteria and none of the exclusion criteria regardless of whether the patients consent to participate or not. This will allow us to evaluate the percentage of patients who accept to participate in the trial and determine the reasons why they do not wish to participate if they refuse'.

As mentioned in the previous submission the percentage of patients taking non-prescribed antibiotics is not negligible in this country, and we do not expect this population to be underrepresented in this trial. When it comes to patients taking unnecessary antibiotics the percentage of patients taking antibiotics prescribed by other clinicians is obviously greater than those who are taking non-prescribed drugs. As also stated in the last submission, clinicians will have to answer a questionnaire with the use of questions and vignettes for assessing their level of confidence in stopping antibiotics in different scenarios including cases depicting patients who have taken non-prescribed antibiotics. Only doctors who are confident to stop unnecessary antibiotics in all these scenarios will be invited to participate. The group of patients who initiated an antibiotic course themselves is a key target of the trial, and we do not think this group will be underrepresented.

2. How will the authors account for variability in experience, confidence and comfort levels of GPs in their assessment that "the GP considers that antibiotics are not needed to be taken"? It seems possible that some GPs in your study could have abnormally low or high confidence in their ability to assess whether antibiotics are necessary, this could skew the results significantly, and potentially lead to a greater number of adverse events. Perhaps inviting GPs who have a midrange level of experience (e.g. have been in practice for 5-15 years) might help to equalize their confidence levels? Another strategy would be to administer a questionnaire to the GPs using vignettes or questions to assess their level of confidence in stopping antibiotics, which might make your results more interpretable.

The authors have taken this concern into account and decreased the number of centers involved from 10 to 20. This will help somewhat to reduce heterogeneity, though heterogeneity of the study population and the clinicians may still affect the interpretability of the results.

Response: We appreciate this remark as our goal is to minimise the heterogeneity of the study because this is an open-label clinical trial. Considering this comment, we will consider the inclusion of doctors with 15 to 25 years of experience with the aim of reducing this heterogeneity as much as possible.

3. The authors propose secondary study objectives to study the incidence adverse events (page 9, objective 1) and number of complications within the first 3 months (page 9, objective 4). Though these are not primary objectives, I believe the study will be underpowered to detect most adverse events and complications, and this limitation should be recognized more clearly.
The likelihood that the study will be underpowered has been addressed by the authors as a weakness. The sample size of 240 per group is justified, but I am still concerned that other outcomes may not be assessable with this relatively small sample.
Response: The calculation of the sample size of the trial is based on the main outcome, and as mentioned in the last submission, we do not expect a high number of complications and adverse effects. We have acknowledged this limitation, and it is included as a weakness of the trial.

4. The authors propose additional qualitative secondary study objectives to assess antibiotic consumption (page 9, objective 2) and assess satisfaction (page 9, objective 3). However, the tools used for these assessments are not provided, and the authors do not adequately acknowledge the potential for information bias when relying on patient self-report of these measures.
The questionnaire has now been included, and the authors note that it has been validated. The authors account for 15% loss of participants due to diary non-completion. A word of caution that those participants not completing the diary will likely differ from those who do complete the diary. I recommend analyzing the demographics and other data from these participants who do not complete the diary (whom I understand will not be included in the analysis) to determine if there are significant differences between these participants and others that could introduce bias.
Response: We have added this information to the new version. Patients who do not return the symptom diaries will not be included in the analysis, but the baseline data, which also includes the demographic variables of these patients, will be compared to those who return the diaries to check if the two populations differ. We have added this information to the new version, as follows: 'A bivariate analysis of the baseline data will be performed between patients returning and not returning symptoms diaries for assessing if the latter population differs from the patients included in the study analysis.'
When it comes to antibiotic consumption this information will be collected from the information given in the diaries and checked by the information provided by the Pharmacy Services of the different Health Services. We have also added this information to the new protocol. As far as patient satisfaction and patient belief in the effectiveness of antibiotics are concerned, this information can only be collected in the same symptom diaries.

5. I am concerned that the authors predict a 20% proportion of enrollees will be lost to follow-up (page 10, lines 53-55). This is a high proportion lost, and these patients are likely to be significantly different to those retained in the study, including their compliance with antibiotic recommendations. Loss of this high proportion of participants could significantly affect trial results.
See my comments in #4, above. Though the investigators have decreased the proportion lost from 20% → 15% (for unclear reasons, though a prior study is now cited), I still worry that this could be a source of bias.
Response: We wish to apologise for this misunderstanding. We have now considered the percentage of 15%, because of the percentage of patients who did not return the symptom diaries in a previous trial on delayed prescribing of antibiotics (reference 39). We found this percentage of diary return inferior to other studies which do not oblige patients to return the diaries to the healthcare centre (sending the diary by post for instance). As mentioned in the protocol, clinicians participating in the trial will be obliged to ask patients to return the diaries at the healthcare centre. In addition, the trial

team will call patients who do not return the diaries in order to make this percentage as low as possible, as also mentioned in the protocol. We acknowledge that this percentage of possible losses has been used in other studies, but in order to have avoid a heterogeneous population we will only analyse patients who return their diaries, and therefore an effort will be made to minimise the percentage of losses in all the sites.

6. The authors plan to invite 40 clinicians from 20 different healthcare settings to join their study, anticipating that each clinician will recruit 10-12 participants over the 2-year study period to achieve their projected enrollment of 430 total participants. Though it is a noble goal to include clinicians and patients from diverse practice settings, I am concerned that it will not be feasible or practical to ensure that 40 clinicians can be trained and achieve compliance with the study protocol and recommendations, and retain these skills over a 2-year period while enrolling participants infrequently. Moreover, recruitment of 5-6 participants annually (1 every 2 months) represents a very small proportion of all patients seen in each provider's clinical practice. If, as the authors state, 15% of all patient visits are for viral conditions targeted by this trial, I would estimate that most clinicians would see approximately 1-3 patients every day with viral illnesses that might qualify them for this study. The expectation that each clinician would recruit such a small proportion of the eligible patient population concerns me that clinicians would 'cherry-pick' patients they wished to enroll, choosing to approach potential research subjects not at random, but rather in a way that is convenient to their clinical practice. In my opinion, this has great potential to introduce systematic bias into the sample, and invalidate the results. I think the trial would be more valid if it were restricted to a smaller number of clinicians (~10), and these clinicians were mandated to approach all eligible patients sequentially for enrollment, until they reached their target enrollment of 43 participants each. In addition, I would request the authors specify what data can be collected from participants who chose not to enroll, important information to help future readers decide how applicable trial results are to their own population.
This concern has been partially addressed. There will be a fewer number of centers included with a larger number of participants at each center. However, the number enrolled still represent a small proportion of those seen at each center. I would advocate for decreasing the number of centers involved further, in order to have more participants seen by fewer clinicians. However, this is clearly a tradeoff between consistency and generalizability, and the investigators are in the best position to determine what will be most informative.
Response: Thank you very much for your comments. Considering only a total of 15 investigators each investigator will have to recruit a total of 32 patients. However, the number of patients who meet the inclusion criteria has decreased as only those patients who have taken antibiotics for less than three days will be invited to participate.

New minor points:
1.Page 1, abstract introduction, lines 9-10: I think the rationale that GPs don't stop antibiotics because they are not sure doing so is safe is a weak one. I think the stronger reasons for failure to stop antibiotics in the face of a viral illness are pressures from patients to continue giving some form of therapy and lack of confidence that an illness is truly viral, without any bacterial component. I think this is worth rewording in the abstract and the protocol to clearly state the rationale.
Response: We fully agree with the reviewer. As also mentioned in the third minor point, we have reworded the first paragraph on page 6 and we have summarised this point in the Introduction of the abstract as suggested by the reviewer.

2. Background, page 5, lines 41-49: There is new information here, including a discussion of sort-course antibiotics. It would be helpful to cite evidence that GPs are not using short courses for pneumonia (which are becoming more common in the US). Also, the analogy between UTI and other types of infection isn't very helpful, since the pathophysiology of each infection and each anatomic site is markedly different.

Response: Thank you for the comment. We have changed this accordingly, as follows: ‘The use of shorter therapies, defined as the taking of an antibiotic for 5 days or less, is commonly used for uncomplicated urinary tract infections. Short course regimens – from 3 to 5 days – have also shown to be as effective as longer therapies in community-acquired pneumonia, acute bacterial sinusitis and acute exacerbations of chronic obstructive pulmonary disease but are seldom used by GPs [13]. When it comes to pneumonia, despite the efforts of infectious committees and guidelines developed by different societies, the duration of antibiotic use is still a major issue for which there is a lack of adherence both in primary and secondary care worldwide [14].’
14. Viasus D, Vecino-Moreno M, De La Hoz JM, Carratalà J. Antibiotic stewardship in community-acquired pneumonia. Expert Rev Anti Infect Ther. 2017;15:351-9.
Furthermore, the latest local primary care guidelines still recommend long antibiotic therapy courses ranging from 7 to 14 days [semFYC guideline, 2010].

3. Please revise lines 18-20 on page 6 “nobody wants to be seen” to make this a less casual and more formal statement. This manuscript would benefit from a review by a native English speaker with expertise in scientific writing.
Response: An English native scientific writer has revised the manuscript. We have reworded this paragraph as follows: ‘Some studies have also shown that other issues such as uncertainty about diagnosis, ease of follow-up and fear of consequences of non-prescribing, as well as perceived pressure to prescribe and potential conflict with patients which might lead to consequences for the future doctor–patient relationship are more of a concern for GPs continuing to prescribe antibiotics than antibiotic resistance [18]. When it comes to acute infections, GPs might feel uncomfortable to discontinue antibiotic therapy from a patient who subsequently deteriorates, especially if the patient needs to be admitted to hospital.’
18. Tonkin-Crine S, Yardley L, Little P. Antibiotic prescribing for acute respiratory tract infections in primary care: a systematic review and meta-ethnography. J Antimicrob Chemother 2011;66:2215–23.

4. Page 6, lines 27-34: please see my comment #1 above. I think it is important to note that the physician-patient dynamic also places pressure on doctors to continue antibiotics, in order to be seen as ‘doing something’ for their patient.
Response: Thank you for the comment. This statement is now clearer with the inclusion of the previous paragraph.

5. Page 11, line 45: the ‘basal’ visit should be changed to ‘baseline’ visit, I think.
Response: Changed.

We hope now the paper has suitably improved to be accepted for publication in this journal.

VERSION 3 – REVIEW

REVIEWER	Lisa Bebell Massachusetts General Hospital, USA
REVIEW RETURNED	05-Apr-2017

GENERAL COMMENTS	My original major and minor points are described in detail below, and following those points are my assessment of whether they have been addressed. Major Points: 1. The authors propose additional qualitative secondary study objectives to assess antibiotic consumption (page 9, objective 2) and assess satisfaction (page 9, objective 3). However, the tools used for these assessments are not provided, and the authors do not adequately acknowledge the potential for information bias when

	relying on patient self-report of these measures. The questionnaire has now been included, and the authors note that it has been validated. The authors account for 15% loss of participants due to diary non-completion. A word of caution that those participants not completing the diary will likely differ from those who do complete the diary. I recommend analyzing the demographics and other data from these participants who do not complete the diary (whom I understand will not be included in the analysis) to determine if there are significant differences between these participants and others that could introduce bias. This is now noted on page 13, and researchers will compare those returning symptom diaries to those who do not. 2. I am concerned that the authors predict a 20% proportion of enrollees will be lost to follow-up (page 10, lines 53-55). This is a high proportion lost, and these patients are likely to be significantly different to those retained in the study, including their compliance with antibiotic recommendations. Loss of this high proportion of participants could significantly affect trial results. See my comments in #4, above. Though the investigators have decreased the proportion lost from 20% → 15% (for unclear reasons, though a prior study is now cited), I still worry that this could be a source of bias. The authors have addressed this in the new version. 3. The authors plan to invite 40 clinicians from 20 different healthcare settings to join their study, anticipating that each clinician will recruit 10-12 participants over the 2-year study period to achieve their projected enrollment of 430 total participants. Though it is a noble goal to include clinicians and patients from diverse practice settings, I am concerned that it will not be feasible or practical to ensure that 40 clinicians can be trained and achieve compliance with the study protocol and recommendations, and retain these skills over a 2-year period while enrolling participants infrequently. Moreover, recruitment of 5-6 participants annually (1 every 2 months) represents a very small proportion of all patients seen in each provider's clinical practice. If, as the authors state, 15% of all patient visits are for viral conditions targeted by this trial, I would estimate that most clinicians would see approximately 1-3 patients every day with viral illnesses that might qualify them for this study. The expectation that each clinician would recruit such a small proportion of the eligible patient population concerns me that clinicians would 'cherry-pick' patients they wished to enroll, choosing to approach potential research subjects not at random, but rather in a way that is convenient to their clinical practice. In my opinion, this has great potential to introduce systematic bias into the sample, and invalidate the results. I think the trial would be more valid if it were restricted to a smaller number of clinicians (~10), and these clinicians were mandated to approach all eligible patients sequentially for enrollment, until they reached their target enrollment of 43 participants each. In addition, I would request the authors specify what data can be collected from participants who chose not to enroll, important information to help future readers decide how applicable trial results are to their own population. This concern has been partially addressed. There will be a fewer number of centers included with a larger number of participants at

	each center. However, the number enrolled still represent a small proportion of those seen at each center. I would advocate for decreasing the number of centers involved further, in order to have more participants seen by fewer clinicians. However, this is clearly a tradeoff between consistency and generalizability, and the investigators are in the best position to determine what will be most informative. The authors have chosen to remain with the same number of centers. Minor Points: 1. Page 1, abstract introduction, lines 9-10: I think the rationale that GPs don't stop antibiotics because they are not sure doing so is safe is a weak one. I think the stronger reasons for failure to stop antibiotics in the face of a viral illness are pressures from patients to continue giving some form of therapy and lack of confidence that an illness is truly viral, without any bacterial component. I think this is worth rewording in the abstract and the protocol to clearly state the rationale. -Largely addressed. 2. Background, page 5, lines 41-49: There is new information here, including a discussion of sort-course antibiotics. It would be helpful to cite evidence that GPs are not using short courses for pneumonia (which are becoming more common in the US). Also, the analogy between UTI and other types of infection isn't very helpful, since the pathophysiology of each infection and each anatomic site is markedly different. -Largely addressed 3. Please revise lines 18-20 on page 6 "nobody wants to be seen" to make this a less casual and more formal statement. This manuscript would benefit from a review by a native English speaker with expertise in scientific writing. -Addressed. 4. Page 6, lines 27-34: please see my comment #1 above. I think it is important to note that the physician-patient dynamic also places pressure on doctors to continue antibiotics, in order to be seen as 'doing something' for their patient. -Addressed. 5. Page 11, line 45: the 'basal' visit should be changed to 'baseline' visit, I think. -Addressed.

REVIEWER	Jesse Berlin Johnson & Johnson I am a full time employee of Johnson & Johnson. Although we do market antibiotics, I don't perceive any direct conflict in reviewing this protocol. We are fully supportive of appropriate use.
REVIEW RETURNED	08-Apr-2017

GENERAL COMMENTS	I have no further comments. I thank the authors for their detailed responses to earlier concerns.