Author's response to reviews

Title: Sample size considerations using mathematical models: an example with Chlamydia trachomatis infection and its sequelae pelvic inflammatory disease

Authors:

Sereina A Herzog (herzog.sereina@gmail.com)
Andrea Berghold (andrea.berghold@medunigraz.at)
Nicola Low (low@ispm.unibe.ch)

Version: 3
Date: 30 March 2015

Author's response to reviews: see over
Dear Ms Ramos

Thank you for giving us the opportunity to submit a revised version of our manuscript “Sample size considerations using mathematical models: an example with Chlamydia trachomatis infection and its sequelae pelvic inflammatory disease”. Please find below our point-by-point response, which is written in blue. We highlighted in the manuscript our changes due to reviewer comments in blue; changes to ensure that the written English is clear and concise are not highlighted.

Responses to Editor:

Comment 1: All reviewers see merit in the work described in this manuscript however revisions are required. In particular, the authors should provide more clarity on the choice of model type, the choice of parameters and the biological implications of these assumptions, and the methods used to validate the model. The authors should also ensure that the written English is clear and concise. While reviewer 3 does not specify any major compulsory revisions, I feel his suggestion to present a sample size calculation to exemplify the methodology is constructive and would improve the paper.

Response: We edited the writing in our manuscript and the structure within paragraphs due to comments but also in more general terms. We changed, for example, the title of Figure 2 and re-arranged the content of the figure legends of Figure 2-5 (see also response to Reviewer #2, comment 5).

In comment 2 from Reviewer #2 we justify our model choice. We clarified in comment 3 from Reviewer #2 that we did not try to obtain any parameter estimations. We comment on the biological implications and generalizability of our results in response to Reviewer 1, comment 3.

We included an explicit example about sample size calculation as suggested by Reviewer #3 (see response to Reviewer #3, comment 2).

Responses to Referee 1:

Major compulsory revisions

Comment 1: In the supplementary material, the steady state prevalence of women with PID is equal to $s_{2}=p(\gamma / (r+\gamma))$. So is not the fraction of infected women who develop PID in the control group, $f$ (defined this way in lines 159--160), equal to $\gamma / (r+\gamma)$ in $s_{3}$? Please explain in more detail.

Response: Not exactly, the steady state prevalence of women with PID equals the prevalence of chlamydia times the fraction of infected women who will develop PID.

We defined fraction $f$ to be the proportion of all women who get infected which will develop PID, i.e. $f = \frac{\gamma}{\gamma+r}$ for the constant progression type. From this definition it follows that $\gamma = \frac{fr}{1-f}$ and therefore

$$i_2 = \frac{\gamma}{\gamma+r} = pf$$

We extended the equations for the steady state of the control group in the Additional File 1 (Section 1, p2):

“The control group starts at steady state in the absence of the intervention with $s=1-p$, $i_1=p \frac{r}{r(\gamma+r)}=p(1-f)$, and $i_2=p \gamma/(\gamma+r)=pf$.”

Note, the notation for the initial conditions $s_1$, $s_2$, and $s_3$ has been changed to $s$, $i_1$, and $i_2$ due to Reviewer #1, comment 9.
Comment 2: Check that your interpretation of $\gamma = fr/(1-f)$ is correct. I believe setting $\gamma = fr/(1-f)$ will ensure that a desired proportion of the control group will develop $PID$ by the end of the one year follow up, but I am not sure if it ensures the same duration of infection across progression types. Is it the basis of your argument above the assumption that duration of infection is equal to $\gamma + r$ in the constant progression scenario? Please check this assumption is correct mathematically and demonstrate it in the appendix.

Response: Thank you. We changed in the sentence “If PID develops at a constant rate the PID incidence equals $\gamma I_1$; to achieve the same cumulative PID incidence as the other two types of progression we set $\gamma = \frac{fr}{1-f}$. Note, the mean duration of infection is still $1/r$ (see Additional File 1, section 1).” in Methods on p7. We also adapted the legend of Table 1.

We changed the sentence in the Additional File 1 and included a short demonstration about the mean duration of infection for the constant progression (section 1, p2):

“We had to set the progression rate $\gamma = \frac{fr}{1-f}$ to achieve the same cumulative PID incidence in all three types of progression. The duration of infection in the constant progression type is $1/r$:

Total mean duration being infected = mean duration in $I_1$
+ probability going from $I_1$ to $I_2$ times mean duration in $I_2$

= $\frac{1}{r+\gamma} + P[I_1 \text{ to } I_2] \frac{1}{r}$

= $\frac{1}{r+\gamma} + \frac{\gamma}{r+\gamma} \frac{1}{r} = \frac{r+\gamma}{(r+\gamma)\gamma} = \frac{1}{r}$

Comment 3: Please comment further on the biological implications and generalisability of your results. The way I see it, theoretically any values of $s_3$ and $f$ between $0$ and $p$ could be obtained by varying the parameters $r$ and $\gamma$, and using any set of assumptions about progression. Please discuss how your results will hold for situations where $f$ increases in value, or if you relax the constraint $\gamma = fr/(1-f)$ in the constant progression scenario. Line 242 would be a good place to insert this discussion. To some extent you do this in Figure 2, but it would be good to discuss this more generally. Do you expect the sample size required in all scenarios to continue to decrease as $f$ increases, and if so why? I also advise you to include a discussion of the implications of your results for sample size calculations for other types of infection, framed around $R_0$, because this quantity is intimately related to the proportion infected—typically this equals $1-1/R_0$. I believe this is also true for your model.

Response: The fraction $f$ is the proportion of all women who get infected which will develop PID and therefore limited to be between 0 and 1 and not between 0 and chlamydia prevalence $p$. To be able to compare the three different types of progression we defined in the Method section that in absence of an intervention we want to achieve the same PID incidence an all three types of progression. Therefore we cannot relax the constraint about $\gamma = \frac{fr}{1-f}$. The initial condition for the $I_2$ compartment ($i_2$ in the new notation, $s_3$ in the old notation) at steady state in the control group is $i_2 = \frac{\gamma}{r+\gamma} p = p f$, i.e. $i_2$ can vary between 0 and $p$ as stated by the reviewer (see also response to Reviewer #1, comment 1).

What happens to the sample size needed per group if $f$ is increasing?
We illustrate the answer to this question, as mentioned by the reviewer, to some extend in Figure 3 (constant progression and progression at the end) and in Figure S4 (immediate progression) where we vary the fraction $f$ between 7-13%.

We investigated the question in a more general approach.
We included the detailed information in the Additional File 1 as section 4 “The relation between sample size, relative risk and PID incidence while varying fraction developing PID” and adapted the section numbering.

We included a summary of these findings in Results as a second last paragraph

“The sample size needed per group depends on PID incidence in the control group and the RR. Hence, the observed pattern of sample size needed for different values of the infectious duration and fraction developing PID is a combination of the estimated patterns for PID incidence and RR. We therefore investigated the relation between sample size needed per group, RR, and PID incidence while varying the fraction developing PID. For the hypotheses of immediate progression to PID and progression to PID at the end of chlamydia infection, the RR is independent of the fraction developing PID, i.e. increasing the fraction developing PID increases PID incidence and decreases the sample size needed per group if everything else is kept constant. For constant progression to PID the relationship between sample size needed per group and the fraction developing PID is more complicated because the RR also depends on the fraction developing PID. In this situation, PID incidence and RR need to be investigated together to predict the sample size needed per group while varying the fraction developing PID (see Addition file 1, section 4).” on p12.

Discuss the results framed around the basic reproduction number $R_0$:
In our case it is not possible to discuss our results framed around $R_0$.

The basic reproduction number $R_0$ is the average number of secondary infections generated by a typical case in a fully susceptible population. In our model we cannot derive a formula for $R_0$ because we assume a constant force of infection, i.e. an infected woman does not infect another woman in our system; the new infections in our model are a result from infection source outside the population represented by our model.

As shown in response to comment 2 of Reviewer #1 we know that the mean duration of infection in our $SI IS$ model equals $1/r$. Therefore we can investigate the following differential equation system for a SIS model:

$$\frac{dS[t]}{dt} = -\lambda S[t] + rI[t]$$
$$\frac{dI[t]}{dt} = +\lambda S[t] - rI[t]$$

with $\lambda$ being the constant force of infection and $1/r$ the mean duration of infection. We have a closed population ($\frac{dS[t]}{dt} + \frac{dI[t]}{dt} = \text{constant}$), i.e. $S[t] + I[t] = N$. Substituting $S[t] = N - I[t]$ into the second differential equation, we obtain

$$\frac{dI[t]}{dt} = \lambda(N - I[t]) - rI[t] = \lambda N - (r + \lambda)I[t]$$

Solving $\frac{dI[t]}{dt} = 0$, we see that there is only one steady state for this SIS model with $I^* = \frac{\lambda}{r + \lambda}N$ which is stable because $\frac{d(\lambda N - (r + \lambda)I[t])}{dt} = -(r + \lambda)$ and therefore always $<0$. This means that if $I$ is perturbed from its steady state value $I^*$, it returns to $I^*$. The stable steady state $I^*$ equals 0 only if $\lambda = 0$ because the mean duration of infection $1/r > 0$ and $N > 0$. A force of infection $\lambda = 0$ means that there is no infection in the source outside the population represented by our model.
Minor essential revisions
--------------------------------------
These are spelling or typographical errors.

**Comment 4:** Line 100: Should read "structure BY taking advantage"

**Response:** We have changed the sentence to clarify how we used the fact that chlamydia prevalence differs at baseline on p6 in the Method section:

“We ran the model separately for intervention and control groups, using different starting conditions because chlamydia status at baseline differed in the intervention and in the control group (Additional File 1, section 1).[16]”

**Comment 5:** Line 108: Missing full stop in Herzog et al.

**Response:** We have changed the sentence and kept only the reference number.

**Comment 6:** Line 196: Should read 21,000 not 21'000

**Response:** We have inserted this suggestion.

**Comment 7:** Line 316: Change to "suggested the fraction of women who develop PID to be 10%..."

**Response:** We did not include this suggestion because the reported values are estimations from mathematical modelling studies which analysed the POPI trial.

**Comment 8:** Line 317: Change CrI in second confidence interval to CI.

**Response:** Price et al. used a Bayesian analysis and therefore the authors reported a credibility interval for their estimation and not a confidence interval. We therefore keep the abbreviation CrI for the credibility interval as listed in the list of abbreviations.

Discretionary revisions
--------------------------------------
Suggested amendments to the text that may improve readability.

**Comment 9:** Perhaps the notation $s_{1}$, $s_{2}$, $s_{3}$ in the Supplementary material could be changed to read $s$, $i_{1}$, $i_{2}$. Using the notation $s_{i}$ to refer to a fraction of the infected population could be quite confusing.

**Response:** We changed the notation for the initial conditions $s_{1}$, $s_{2}$, and $s_{3}$ to $s$, $i_{1}$, and $i_{2}$ in the Additional File 1, section 1, p.2.

**Comment 10:** Lines 48-49: suggest reword to "its sequelae can be accounted for in sample size calculations by using mathematical modelling."

**Response:** We have adapted this suggestion by changing the sentence in the Conclusion section of the abstract on p3 to
“Mathematical modelling helps to understand the temporal relationship between an infection and its sequelae and can show how uncertainties about natural history parameters affect sample size calculations when planning a RCT.”

We do not agree deleting “RCT” as a key word in our conclusion as we considered only a RCT study design.

Comment 11: Lines 60-61: long and confusing sentence, consider splitting into two sentences.

Response: We changed the sentence to

“The success of an intervention to prevent the disease complications by reducing exposure to infection is influenced by the natural history of the infection. Different assumptions about the temporal relationship between the infection and the development of sequelae can affect the expected effect size of an intervention. To calculate the required sample size for a trial investigating the reduction in complications of and infection, we need to make assumptions about the effect size of the intervention (e.g. the reduction in relative risk) and the incidence of the infection sequelae.” on p4.

Comment 12: Line 189: Suggest "Scenario 1 utilises the assumption that..."

Response: We apologise if the reviewer was confused. Scenario 1 does not utilise or use the assumption that PID can develop throughout the infection period. The investigators of the POPI trial used in their first sample size calculation the assumptions that the relative risk is 0.48 and a PID incidence of 2%. The relationship between the power of a study and the sample size required per group calculated with the POPI trial assumptions is compatible with the calculations assuming that PID can develop throughout the infection period. We did the calculation for all three types of progression. We changed the sentences in the Results as follows,

“In scenario 1, the relationship between the power of a trial and the sample size required per group is compatible with the hypothesis that PID can develop throughout the course of infection (Figure 2A). Assuming a constant progression rate from chlamydia to PID results in a RR of 0.49, which is close to the original RR assumption in the POPI trial (RR=0.48). If chlamydia progresses to PID only at the end of the infectious period, the model predicts a RR of 0.39.” on p10.

Comment 13: Line 191: Suggest "Scenario 2 models a disease process where PID develops..."

Response: This is a similar misunderstanding as in comment 12 of Reviewer #1. The wording “scenario 2” covers the assumptions used by the POPI trial investigators for the second sample size calculation, i.e. 3% PID incidence and a relative risk of 0.44. The relationship between the power of a study and the sample size required per group calculated with these POPI trial assumptions is now compatible with the calculations assuming that PID can develop at the end of the infection period. We clarified that changing the sentence to

“In scenario 2, the assumptions used for the second sample size calculation (RR=0.44, 3% PID incidence) are more compatible with the hypothesis that PID develops at the end of a chlamydia infection (Figure 2B). This model results in a RR of 0.39, as before. In a model that assumes a constant progression rate from chlamydia to PID, the predicted RR=0.56.” on p10.
Comment 14: Line 242: Suggest "There are optimal follow-up times that minimise the sample size needed per group..."

Response: We have inserted an adapted version of this suggestion on p13: “The last analysis in our study showed that there are optimal follow-up times which minimise the sample size needed per group [...]”

Comment 15: Line 251: suggest "This study showed..."

Response: We have inserted this suggestion on p13.

Comment 16: Line 252: Suggest "RCTs. The implications of our findings were illustrated using the example of chlamydia infection and its sequelae PID by re-"

Response: We have adapted this suggestion on p13. The Discussion and conclusion section starts now with

“This study showed how a mathematical model can be used to inform sample size considerations for RCTs. We used the example of a screening intervention to prevent chlamydia infection and PID as a complication by re-examining published sample size calculations from the POPI trial.”

Comment 17: Line 253: Suggest "Different temporal dependency assumptions regarding chlamydia and PID are reflected in different PID incidence, and hence different RR values used in sample size calculations. Changing the underlying temporal dependency assumption can affect required sample size considerably, even for small changes in..."

Response: The suggested sentences have a different meaning compared to the original sentences. We assume that this is a consequence of the misunderstanding about the description of the results for scenario 1 and 2 (see response to Reviewer #1, comment 12 and comment 13).

We changed the sentences to

"Different sets of assumptions about PID incidence and the RR values used for the sample size calculation in the POPI trial required different hypotheses about the temporal relationship between chlamydia and PID."

on p13.

Comment 18: Line 320: Suggest "implications for the future planning of RCTs."

Response: We have inserted this suggestion on p16.
Responses to Referee 2:

Major Compulsory Revisions:

Comment 1: Throughout, I’d found two main areas where the manuscript could be improved: the written English, and the details of your model choice (this includes why you chose a particular model structure, and how you validated it).

For the first, I found myself having to re-read the “Methods”, “Results”, “Conclusions”, and figure captions multiple times. And as a result, I frequently found myself guessing at what you were trying to communicate.

As a specific example, I found the caption of Figure 2 difficult to follow. In an attempt to understand what you’d written, I provide an edited version (i.e. my translation) of the figure caption:

“Figure 2. Estimated sample sizes under different assumptions for Pelvic Inflammatory Disease (PID) incidence in the Prevention of Pelvic Infection (POPI) study.

Plotted curves represent PID incidence in the study group relative to the control group of the original POPI study (green lines) and for two modelled scenarios: one where PID develops continually (dashed-dotted lines), and one where it develops at the end of infection (dashed lines). Relative risk curves for a third model scenario (where PID immediately follows Chlamydia infection) are not shown. Panels A and B separate our analysis based on the projected follow-up incidence of PID in the original POPI study: 2% per year (Panel A), and 3% per year (Panel B). The red circle represents the sample size with 80% power in the original POPI study.”

While I feel it will benefit your manuscript to have someone edit the writing, I do not expect you to keep the edits I have made to the figure caption. (I only hope they help provide some insight into how I interpreted your writing!)

Response:

Written English and structure

We edited the writing in our manuscript and the structure within paragraphs due to comments but also in more general terms.

- We changed the title of Figure 2 and re-arranged the content of the figure legends of Figure 2-5 (see also response to Reviewer #2, comment 5).

- For example, we have rewritten in the Result section of the abstract on p2

  “The assumed event rates and effect sizes used for the sample size calculation implicitly determined the temporal relationship between chlamydia infection and PID in the model. Even small changes in the assumed PID incidence and relative risk (RR) led to considerable differences in the hypothesised mechanism of PID development. The RR and the sample size needed per group also depend on the natural history parameters of chlamydia.”

  as well as the start of the Result section on p10

  “Different sets of published assumptions about PID incidence rates and the size of intervention effect lead to different conclusions about the temporal relationship between chlamydia infection and PID (Figure 2).”

- We brought the information about the POPI trial together, i.e. we now state

  “Women were randomly allocated to an intervention group, which received immediate testing and treatment for women with positive chlamydia test results. The control group reflected
routine care, but swabs were collected at baseline and stored. Testing and treatment were then deferred for one year.”

in the Background section on p5 where the POPI trial is presented instead of giving this information later in the Method section.

- We adapted slightly the subheadings for our two approaches to investigate sample size calculations: “Sample size calculations used in the POPI trial” and “Generic sample size calculation using the mathematical model”.

- We restructured the Discussion section by shortening the previous second paragraph to one sentence which we added to the first paragraph on p14

  “Not to forget, the RR and the corresponding sample size needed per group depend also on the other natural history parameters of the infection, however, not all infection parameters have the same impact.”

and by adding a paragraph on p15 which discusses all three postulated temporal relationships together with the fraction of women who develop PID after a chlamydial infection instead of discussing only the immediate progression (previous third paragraph of the discussion):

“The POPI trial data and our modelling study together allow an interpretation of the postulated temporal relationship between chlamydial infection and PID in the context of published RCT evidence. The first set of assumptions used to calculate the required sample size (chlamydia prevalence 7%, PID incidence 2%, RR 0.48) was compatible with the hypothesis that PID can occur at any time during the infectious period of C. trachomatis. This mechanism is supported by the effects of similar screening and treatment interventions in other RCTs.[15,20,27] The observed chlamydia prevalence and PID incidence in the POPI trial were close the actual assumptions but the effect size was smaller, so the trial was underpowered. The second set of assumptions was quite similar (chlamydia prevalence 7%, PID incidence 3%, RR 0.44) but the mathematical model showed that these conditions could only be satisfied if C. trachomatis progresses to PID at the very end of the infectious period. This hypothesis lacks biological plausibility, given that chlamydial infectious load in the lower genital tract should be lower at the end than the beginning of infection.[26] The results of published RCTs also argue against immediate progression to PID after C. trachomatis infection.[15,20,27] Our model allowed us to examine the probability of chlamydia infection progressing to PID. In our model, both sets of sample size assumptions would require a high fraction of chlamydia infection progressing to PID (28.6% with yearly PID incidence of 2% and 42.9% PID incidence 3%). Estimates of this size have been used in several cost-effectiveness studies.[7] However, two recent modelling studies analysing the results of the POPI trial estimated that 10% (95% CI 7-13%) respectively 12% (95% CrI 2-24) of women develop PID.[16,28]”

Information for clarification:
We added in the revised manuscript information for clarification.

- “There is great interest in interventions that could reduce the risk of chlamydia-associated PID because this might prevent future tubal factor infertility.” on p5 in the Background section.

- “ […] (RR=risk of PID in intervention group/risk of PID in control group).” on p7 in the Method section to clarify how the relative risk was calculated.
We extended the information about the background why the POPI trial investigators revised their sample size calculation

“In that RCT, 7% of women in the intervention group had a positive chlamydia test result at baseline and the incidence rates of PID one year later were 8 per 10,000 woman months in the intervention and 18 per 10,000 woman months in the control group (RR 0.44, 95% CI 0.20-0.90). During the POPI trial, the investigators revised their sample size calculation, owing to slow enrolment. The revised calculation cited a cohort study suggesting a higher PID incidence (9.5% to 12.0% over four years in three different groups of women).”

on p8 in the Method section.

Model choice:
We justify our model choice in comment 2 from Reviewer #2.

Comment 2: For the second, you don’t really discuss why you focused on a single SIS model. Did you examine your outcomes using different model structures (e.g., and SIRS model)? While I do not think your analysis has to be exhaustive to be informative, I do think it could be a little more extensive. At the very least, you should discuss why you chose one model structure over others.

For example, the concept of immunity has been shown to be important for Chlamydia (Vickers and Osgood, 2010; Johnson et al., 2011), and thus will be important for the development of PID. Your choice of model structure will also impact policy projections (Pitman et al., 2012). It might be interesting to check whether your estimated sample sizes might vary under different assumptions about Chlamydia’s “natural history”.

Response: We focused on a single SIS model and did not examine our outcomes using other model structures until now.

We acknowledge that there is an ongoing discussion about the existence and duration of immunity after a chlamydia infection [Gottlieb JID 2010;201 (suppl 2):S190-S204]. We agree that whether an immunity stage is included in the model structure or not can make a difference in the analysis of reported chlamydia infections and projections about how chlamydia infection is spreading. However, we assumed that a feasible sample size per group for conducting a study is too small to change the population prevalence of chlamydia and therefore we used a constant force of infection in our model.

Due to the Reviewer comment we investigated how our results for the “Sample size calculations used in the POPI trial” would alter if we include an immunity stage. We assumed that there is no immunity after treatment and that women in the immunity stage cannot be identified (i.e. not be differentiated from susceptible women or infected women). We did not consider that PID develops after chlamydia infection period. i.e. we investigated the same three types of progression.

Including an immunity stage does not alter the interpretation of what we observed for scenario 1 and 2, i.e. scenario 1 is still compatible with the assumption that PID develops throughout the infection period (=constant progression) and scenario 2 is still compatible with the assumption that PID develops at the end of the infection. The relative risk (RR) is increasing for all three types of progression with increasing duration of immunity. This means for the constant progression and the progression at the end that the effect size is decreasing and therefore sample size needed per group is increasing. For the immediate progression, an increasing RR results in an increase in the effect size because $RR>1$ and therefore a decrease in the sample size needed per group. Note, a $RR>1$ means that the predicted PID incidence in the intervention group is higher than in the control group.

We extended the paragraph with the limitations of our study
“Finally, we focused on a SIS model but including an immunity stage did not alter the results in our study about the sample size considerations in the POPI trial (Additional File 1, section 7).”

on p14 of the Discussion section.

We added to the ‘Additional File 1’ section 7 "Including an immunity stage - a SIRS model" to report in details the results including an immunity stage as additional analysis.

Reference:

Comment 3: You state that “The force of infection λ is calibrated so that the steady state prevalence in the model is equal to the prevalence p of the study population” (Methods, line 114-115). However, I think it might be a good idea to outline how the model was calibrated (e.g. what optimisation algorithm was used), how well the model “fit” observed data (this might be tricky since it sounds like you calibrated against a single data point), and if you cross checked the model against any other epidemiological signatures that weren’t used in the calibration?

Given the complexity of accurate parameter estimation, calibrating may force re-estimation of uncertain or implausible parameters (Pitman et al., 2012). Did you try to re-optimise the other previously calibrated parameters (such as the duration of infection or the fraction of women who progress to PID)?

Response: Using the phrase “The force of infection λ is calibrated so that …” gave the wrong impression that we fitted the model. We did not try to obtain any parameter estimations using our model. We clarified that using the word ‘calculated’ instead of ‘calibrated’ on p6.

Minor Essential Revisions:

Comment 4: Table 1: Include units for “infection progression”, 1/y. I’m assuming its units are in “days”.

Response: We have inserted this suggestion in Table 1.

Comment 5: Figure Legends: I think you could re-arrange the content of the figure legends to improve their clarity. Please see my example for the caption of Figure 2 above.

Response: Thank you. We changed the title of Figure 2 and re-arranged the content of the figure legends of Figure 2-5 on p20 ff. We had to keep the titles of the figures short with a maximum of 15 words and restrict detailed legends to a maximum of 300 words according to the instructions for authors.

For Figure 2, we included the relative risk values derived by the different progression types in the main text instead of putting them in the figure legend:

“Assuming a constant progression rate from chlamydia to PID results in a RR of 0.49, which is close to the original RR assumption in the POPI trial (RR=0.48). If chlamydia progresses to PID only at the end of the infectious period, the model predicts a RR of 0.39.” on p10 and
“This model results in a RR of 0.39, as before. In a model that assumes a constant progression rate from chlamydia to PID, the predicted RR=0.56.” on p10.

Comment 6: Sample size estimate has an apostrophe (21’000) instead of a comma (21,000). Maybe this was introduced when uploading the manuscript?

Response: We have corrected this on p10.

Comment 7: Additional file: According to Table 1 (in the main text), \(t\) has the units “months”, while \(r, \lambda,\) and \(\gamma\) have the units “days”. Assuming these have remained consistent, the terms \(rt, -(r+\gamma)t,\) and \(-(r+\lambda)t\) (for calculating cumulative incidence) might have mismatched units. Maybe you’d converted one time unit to another “behind-the-scenes”? I’m assuming you did, but it might be worth mentioning that “months were converted to days before calculating \((t)\)”, or vice versa.

Response: We changed in Table 1 the unit for the follow-up time from months to days and adapted therefore also the values.
Responses to Referee 3:

Minor revisions

-----------------------------

Comment 1: In the introduction I think it would be worthwhile outlining what is involved in a sample size calculation—particularly the need to estimate the possible relative risk that can be measured prior to performing the study.

Response: We changed the end of the first paragraph in the Background to

“Different assumptions about the temporal relationship between the infection and the development of sequelae can affect the expected effect size of an intervention. To calculate the required sample size for a trial investigating the reduction in complications of and infection, we need to make assumptions about the effect size of the intervention (e.g. the reduction in relative risk) and the incidence of the infection sequelae.” on p4.

Comment 2: The authors present their approach generally and for a previously completed study. I think it would also be valuable to present a sample size calculation for a theoretical future. This would further demonstrate the importance of this approach when performing sample size calculations in the planning stages of a study and show readers how to incorporate the approach into their work.

Response: We included explicitly the results using the baseline values stated in Table 1 ‘Generic sample size calculation’. We therefore changed in the Methods the start of section “Generic sample size calculation using the mathematical model” to

“First, we used baseline values for all parameters (Table 1) to derive the PID incidence in the control group, the RR for each type of progression and the sample size needed per group. We then changed the duration of infection and [...]” on p9.

We also changed in the Results the start of section “Generic sample size calculation using the mathematical model” to

“Assuming a chlamydia prevalence of 7%, a mean infection duration of one year, and that 10% of infected women will develop PID, we expect a PID incidence of 0.007 after one year of follow-up. Table 2 shows the resulting RR and the sample size needed per group for each type of progression from chlamydia infection to PID. For example, if PID occurs at a constant progression rate, PID incidence is 0.007 per year in the control group and 0.0029 per year in the intervention group. A chi-squared test with a 5% two-sided significance level will have 80% power to detect this RR of 0.42 with a sample size of 4,654 women in each group. With this sample size we would expect 33 PID cases in the control group and 14 PID cases in the intervention group after one year of follow-up.” on p11 and included a Table 2 with the results.

Comment 3: In the discussion it would nice if the authors could describe how this approach could be applied to other diseases and study types.

Response: We changed the last paragraph in the Discussion to describe how our approach can be applied to other infections and diseases

“Our approach can be applied to other infections and diseases. The mathematical model introduced in this study could be adapted to other sexually transmitted infections such as Mycoplasma genitalium, for which there is great uncertainty about the natural history, but for which screening interventions have been advocated.[29] The model can also be extended to incorporate different types of infection (e.g. those for which immunity after infection is more important) and other assumptions about the temporal relationship between the infection and the complication.” on p16.
We now state on p15, where we give examples for other temporal relationships between chlamydia infection and PID, more explicit what the advantage of our approach is and added a further publication “Gray and colleagues assumed a uniform rate of progression from chlamydia infection to PID in their investigation of the effects of a chlamydia vaccine.[26] The advantage of our approach is that we have compared the implications of different assumptions for RCT design and planning.”

How our approach could be applied to study types is more difficult to answer. We examined a two-armed RCT as a study design to investigate the potential influence of a single round screening for chlamydia on PID incidence. We are aware that not all types of interventions can be investigated using RCT design due to e.g. ethical reasons. In such cases observational studies are often the only possibility. We plan to conduct future modelling studies that investigate the use of mathematical models for prospective planned observational studies. However, it is too early to speculate how the mathematical models would look like and we therefore did not include this aspect of your suggestion.

Comment 4: I noticed a typo in the caption of Table 1: remove (max 15 words) from the end.

Response: Thank you. We removed ‘(max 15 words)’ from the caption of Table 1.