Manuscript Number: **EPIDEMICS-D-24-00069**    
  
Manuscript title: **Estimating nosocomial transmission of micro-organisms in hospital settings using patient records and culture data**  
Dear Mr Cascante Vega,  
  
Thank you for submitting your manuscript to Epidemics.  
  
We have completed the evaluation of your manuscript, which has been seen by 3 reviewers and a editor with expertise in the field. The reviewers and editor recognize that, while the study does not perhaps provide a lot of new insights on nosocomial transmission, the modeling approach is well described and of interest to Epidemics. They have made several comments and suggestions for major revision. We invite you to resubmit your manuscript after addressing the comments below. Please resubmit your revised manuscript by **Sep 10, 2024**.  
  
When revising your manuscript, please consider all issues mentioned in the reviewers' comments carefully: please outline every change made in response to their comments and provide suitable rebuttals for any comments not addressed. Please note that your revised submission may need to be re-reviewed.   
  
To submit your revised manuscript, please log in as an author at https://www.editorialmanager.com/epidemics/, and navigate to the "Submissions Needing Revision" folder.    
  
Research Elements (optional)  
This journal encourages you to share research objects - including your raw data, methods, protocols, software, hardware and more – which support your original research article in a Research Elements journal. Research Elements are open access, multidisciplinary, peer-reviewed journals which make the objects associated with your research more discoverable, trustworthy and promote replicability and reproducibility. As open access journals, there may be an Article Publishing Charge if your paper is accepted for publication. Find out more about the Research Elements journals at https://www.elsevier.com/authors/tools-and-resources/research-elements-journals?dgcid=ec\_em\_research\_elements\_email.  
  
  
Epidemics values your contribution and I look forward to receiving your revised manuscript.  
  
Kind regards,

Pierre-Yves Boelle, Editor

Cecile Viboud, Editor-in-Chief    
  
Epidemics  
  
**Editor's comments**  
**Editor 1**

I concur with the reviewers that bypassing HCWs in such modelling is quite a strong assumption... How to interpretet the estimated transmission parameters in a hospital hygiene perspective is therefore difficult. At best, the model would be changed, but at least, some discussion regarding the consequences is warranted.

Answer. Thank you for the comment. We agree that our inference of nosocomial transmission rate embeds substantial uncertainty as it models person-to-person transmission, vectored transmission by HCWs, and also fomite transmission for example by contact with infected devices with the same parameter. However, in the absence of further data to parametrize the movement of HCWs or carriage data of possible environmental reservoirs we limit ourselves to the patient movement data available. In the section *Understanding nosocomial transmission,* we discuss these limitations and suggest that future models should explicitly include those modes of transmission.

To further discuss the consequences we extend that section in the discussion: Fomite transmission could contribute to the nosocomial acquired carriage by both contact with patients for example by manipulation of uncleaned devices, or by helping as reservoirs for vectorial transmission by HCWs. If those modes of transmission contribute linearly to the estimated nosocomial transmission rate, then a fraction of our estimate would be attributed to that specific mode. If the contribution is non-linear, then it would be necessary to investigate if that interaction is frequency and thus depends on the fraction of HCWs colonized on a ward at a given time or density-dependent. The results of such interactions in nosocomial transmission would be harder to resolve without dynamic measurements of environmental and HCW colonization. However, the non-linear nature of the interaction enriches dynamical outcomes and should used to inform better data collection to parametrize model-inference systems. Without additional data streams without carefully selected data disentangling the contribution of other indirect modes of transmission is challenging.

**Editor 2**

A second point is I did not understand how the ABM formulation was essential to your results? It seems that the transmission model is Markovian and could have been simulated otherwise. In other words, you could have simulated the hospital with standard compartmental models, and I don't think that the reslts would be changed. It would be good to show whether the ABM is really of importance, or if it is just a choice for simulation.

Answer. Thanks for the comment as it allows us to further describe our model inference approach. The preference of an ABM versus a compartmental stochastic model is described below.

1. Both the observational model, as well as the process model, use individual-level test data and movement data. Thus, given the test data on our simulations, the same agents are always predefined to be tested. This approach allows to have a dynamical observational model as well as parametrizes an individual-level testing sensitivity. As the probability of testing depends on the patient's status we think that the results are likely to change.
2. During the study period, few wards are consistently populated, and thus using for example a stochastic compartmental model requires aggregation of the wards to a coarser spatial scale. This also requires a different parametrization of the observational model, for example by assuming a Binomial observational model with the number of trials equal to the number of tests and the probability of success proportional to the fraction of carriers inside a spatial unit.
3. Finally, the movement of patients in the ABM is as defined by the individual-level data replaying what happened and was reported in the hospital. Using a different modeling approach that rather models a coarser aggregation than the individual level would require to design an additional movement model.

**Reviewers' comments**  
**Reviewer #1**  
  
**Comment 1.** A justification is missing in the methods section or supplementary material for the assumption of a Gaussian observation process.  
Answer. Thanks for this comment. We don't assume a Gaussian observational process. The method used for data assimilation, the ensemble adjustment Kalman filter (EAKF) models both the likelihood as Gaussians so the posterior is also a Gaussian. However, our observational process directly observes the patient status in each ward so it is not assumed that the number of detected carriers is drawn from a normal distribution.

**Comment 2.** The observed data is assumed to be normally distributed, should this not be a lognormal distribution? I assume you are constraining to positive values.

Answer. Thanks for the comment. The EAKF assumes the observed/assimilated data is normally distributed so the posterior can be updated with a deterministic mathematical formula. This way we only parametrize the observed data using an observational error variance (OEV, as indicated in the Supplementary Information section **Simulation-based inference framework**) that we model proportional to the data, that is to the number of positive cultures. There's no need to constrain it to positive values as the OEV enters the update deterministically. It is possible, however, that the posterior of both the parameters and the observations is outside the prior range or negative. In the process model updates, we constrain the parameters inside the specified prior range and the observations to be positive.  
  
**Comment 3.** Figures are missing axis labels and alphabetic labels (e.g. A, B which are used in the figure caption) throughout the manuscript and supplementary materials. For example, in Figure 1 and Figure 4. Please add these to ensure the figures are clear.

Answer. Thanks for this. We might forget to add the labels in the last update of the manuscript. In the current submission, we included them. DO THIS.  
  
**Comment 4.** Figure 4 caption refers to a legend which isn’t present. Please add and check other figures.

Answer. Thanks  
  
**Comment 5.** Figure S12 refers to ‘institutions’, this terminology is not used anywhere else in the manuscript. The distinction between institution and hospital is not clear to me, please clarify.

Answer. Thank for making us notice this, it is a slip. We changed it to 'hospital'.

**Comment 6.** The abstract would benefit from stating that an ensemble adjustment Kalman filter is used.

Answer. Thanks,

**Comment 7.** A rationale is needed for using the ensemble adjustment Kalman filter method over a method more often used for inference of transmission rate parameters, such as ABC/SMC/MCMC.

Answer.

**Comment 8.** The study period coinciding with peak COVID-19 cases in the city is mentioned as a limitation. Further discussion on how this might affect the study results would be beneficial. Did you consider fitting multiple transmission rates for different time periods, perhaps when different levels of restrictions were in places?

Answer. Thanks.  
  
**Comment 9.** A more detailed discussion is needed on the types of infectious pressure that the ABM model is not capturing, particularly the lack of hospital staff in the model. These are briefly mentioned in Understanding nosocomial transmission. Ideally when modelling nosocomial transmission an ABM would either explicitly model HCWs and hospital staff as agents in the model, or implicitly include them within the force of infection. Moreover, infection dynamics could be disentangled further if other sources were considered, such as: staff infectious pressure, spatial infectious pressure from other wards (patients/staff) that may be connected, and a background infection pressure. I assume staff data was unavailable for this study?

Answer. Thanks for this comment as it allows us to further strengthen our discussion around the estimated parameter. As the reviewer indicates we limit ourselves to delineating contacts of just the patients as the Health Care Worker (HCW) data was unavailable. This comment is in nature similar to the comment of **Editor 1** which the reviewer can see at the beginning of the document. We extended our discussion in the section *Understanding nosocomial transmission* to further clarify how our nosocomial transmission rate embeds both, person-to-person, and vectored transmission via HCWs and possible transmission by contact via environmental reservoirs. We also suggest what are the possible contributions of those modes to our nosocomial transmission rate depending on the nature of the contribution.  
  
**Comment 10.** It is unclear to me why the force of infection for each ward only considers the infection/colonization status of patients in the same ward. Did you consider adding a ‘background’ transmission rate parameter which scaled with the number of infections in the hospital at that time?

Answer. Thanks for this comment, it sounds like the reviewer suggested that background transmission might model for example fomite transmission that could be represented as proportional to the prevalence in the hospital. Our first inference approach attempted to estimate two transmission rates, one by explicit modeling of environmental reservoirs which contributed to the force of infection via an additional environmental transmission rate parameter. We were unable to accurately identify both parameters using the same data we presented. As a consequence, we decided to only infer one transmission rate at the expense of embedding multiple modes of transmission. The data used to inform the inference is so sparse and represents such a small fraction of the prevalence that without additional data any other parametrization is another assumption. This in addition to the range of pathogenic bacteria used.  
  
**Comment 11.** A daily contact network was used rather than a finer-scale (e.g. hourly) to capture patient movements, was this for computational benefit? It is not clear to me if a patient was admitted into A&E and subsequently to an inpatient ward on the same day, whether they would only appear in the inpatient ward of the contact network? If so, the model may not be capturing transmission in these staging areas.

Answer. We decided to use daily hours as other modes of transmission were not explicitly modeled. Consider for example, that a carrier inpatient is admitted to a ward for a few hours of a day and contributes to the environmental reservoir; later that day

**Comment 12.** The final three sentences of the introduction seem unnecessary and should be covered in the abstract/results and discussion.

Answer. Thanks.  
  
**Reviewer #2**

This is an interesting paper and the methods are well explained. However, by focussing on carriage, particularly in the case of *E. coli*, I worry that the conclusions are not helpful or informative and this needs to be addressed and discussed in more detail. The model does not distinguish between commensal bacteria and pathogenic or resistant bacteria and the message that *E. coli* is not commonly transmitted in hospital settings is likely not true, and is in fact an artefact of the model structure that only allows for a single colonization event with *E. coli*. While all models are abstractions, this issue cannot be ignored.  
  
Here is a list of comments, please also see annotations in the PDF document attached (if you can't get the attached file, inquire with the journal).

**Comment 1.** There is no definition of what is classed as a nosocomial case in the underlying data. Many cases that are considered community acquired based on days in hospital before infection is detected are in fact potentially related to a previous hospital event. Is this included in the definition of nosocomial used here?  
  
**Comment 2.** Are some sites more likely to be sampled than others, therefore effecting the estimated prevalence in the underlying data?  
  
**Comment 3.** The parameters underlying the model are drawn from literature outside of the region being simulated. While I appreciate that this is a limitation that cannot be avoided when there are no available data, a full sensitivity analysis should be performed on these parameters to determine the impact that they have. For example some of the E coli estimates for community carriage are very high and this reduces the impact of nosocomial transmission in this model framework where patients already carrying E coli cannot be infected again. A sensitivity analysis is essential here as it is highly unlikely these results hold true for antibiotic resistant E coli which likely have a much lower prevalence in the community. I also feel that this point should be drawn out in the discussion. The introduction talks about ABRO, but this is a very different problem to what is being modelled in this paper.  
  
**Comment 4.** limitations could be addressed by a global sensitivity analysis as suggested.  
  
**Comment 5.** The layout of the manuscript is excellent and the subheadings are very clear and make the manuscript very easy to navigate. However the writing in this manuscript feels rushed and is confusing in places. The manuscript would benefit from a thorough readthrough and edit by someone less familiar with it's content.  
  
**Reviewer #3**

I feel the article applies interesting methodology and a lot of technical interesting work has been performed, but still, I wonder how much added value the manuscript has to the existing literature. The biological conclusions, that E. coli spreads not very well and MRSA reasonably well in hospital environments is not new and I do not trust the numerical values too much.  
  
**Comment 1.** My impression is that fixing the importation rate is quite a big assumption. Especially when the amount of data is limited, as is the case as primarily infection data are being used, differentiation between importation and transmission is difficult as both are positively related to a high prevalence. A choice of the importation rate based on the literature, from many different countries, can have quite some impact on the transmission parameter as is seen in the sensitivity analysis.  
  
**Comment 2.** Also, I find the assumption of a constant transmission parameter in all wards and all hospitals rather unrealistic? I expect that there will be at least a three-fold difference between wards (with ICUs among the wards with the highest transmission parameters). How should we interpret the average transmission parameter which is calculated in this study?  
  
**Comment 3.** Why are the analyses to determine identifiability of the parameters only done with importation rates of 25% and 50%? A lower value of 5% or 10% seems to reflect the situation for MRSA and Pseudomonas better.  
  
**Comment 4.** I find the choice of the time period a bit strange, a relatively short time period is chosen to ensure that the admission prevalence is more or less constant, but the time period does overlaps with the COVID-19 epidemic which had a big impact on processes in the hospital. Moreover, from Figure 1 I conclude that the number of hospitalized patient is zero at the start of the study? It seems that only data are used from the starting date onwards, but this means that patients already present in the hospital at the starting data are not part of the estimation process, which means that in the first weeks/months, the prevalence will be relatively low. This can also be seen form the empirical colonization data. I think only the data should be used from the moment the hospital size is in its stable distribution as the initial data do not represent a stable situation.

**Reviewer's Responses to Questions**

Note: In order to effectively convey your recommendations for improvement to the author(s), and help editors make well-informed and efficient decisions, we ask you to answer the following specific questions about the manuscript and provide additional suggestions where appropriate.  
  
1. Are the objectives and the rationale of the study clearly stated?  
  
Please provide suggestions to the author(s) on how to improve the clarity of the objectives and rationale of the study. Please number each suggestion so that author(s) can more easily respond.

Reviewer #1: Yes, these are clearly stated.

Reviewer #2: Yes, the study aims to quantify the nosocomial transmission rates of key pathogens and determine their effective detection rate.

Reviewer #3: Yes. Although the assumptions could be stated a bit earlier. Mentioning in the introduction that a frequency-dependent transmission model per ward with the same transmission parameter per ward is used would be helpful to understand the model.

2. If applicable, is the application/theory/method/study reported in sufficient detail to allow for its replicability and/or reproducibility?  
  
Please provide suggestions to the author(s) on how to improve the replicability/reproducibility of their study. Please number each suggestion so that the author(s) can more easily respond.

Reviewer #1: Mark as appropriate with an X:  
Yes [X] No [] N/A []  
Provide further comments here:  
The supplementary materials provided detailed explanations of the model and the model code is available.

Reviewer #2: Mark as appropriate with an X:  
Yes [X] No [] N/A []  
Provide further comments here:

Reviewer #3: Mark as appropriate with an X:  
Yes [X] No [] N/A []  
Provide further comments here:

3. If applicable, are statistical analyses, controls, sampling mechanism, and statistical reporting (e.g., P-values, CIs, effect sizes) appropriate and well described?  
  
Please clearly indicate if the manuscript requires additional peer review by a statistician. Kindly provide suggestions to the author(s) on how to improve the statistical analyses, controls, sampling mechanism, or statistical reporting. Please number each suggestion so that the author(s) can more easily respond.

Reviewer #1

Mark as appropriate with an X:  
Yes [X] No [] N/A []  
Provide further comments here:  
Yes, a few minor points:  
  
1) A justification is missing in the methods section or supplementary material for the assumption of a Gaussian observation process.  
  
2) The observed data is assumed to be normally distributed, should this not be a lognormal distribution? I assume you are constraining to positive values.

Reviewer #2: Mark as appropriate with an X:  
Yes [] No [X] N/A []  
Provide further comments here:  
  
There is no definition of what is classed as a nosocomial case in the underlying data. Many cases that are considered community acquired based on days in hospital before infection is detected are in fact potentially related to a previous hospital event. Is this included in the definition of nosocomial used here?  
  
Are some sites more likely to be sampled than others, therefore effecting the estimated prevalence in the underlying data?  
  
The parameters underlying the model are drawn from literature outside of the region being simulated. While I appreciate that this is a limitation that cannot be avoided when there are no available data, a full sensitivity analysis should be performed on these parameters to determine the impact that they have. For example some of the E coli estimates for community carriage are very high and this reduces the impact of nosocomial transmission in this model framework where patients already carrying E coli cannot be infected again. A sensitivity analysis is essential here as it is highly unlikely these results hold true for antibiotic resistant E coli which likely have a much lower prevalence in the community. I also feel that this point should be drawn out in the discussion. The introduction talks about ABRO, but this is a very different problem to what is being modelled in this paper.

Reviewer #3: Mark as appropriate with an X:  
Yes [] No [] N/A []  
Provide further comments here:

4. Could the manuscript benefit from additional tables or figures, or from improving or removing (some of the) existing ones?  
  
Please provide specific suggestions for improvements, removals, or additions of figures or tables. Please number each suggestion so that author(s) can more easily respond.

Reviewer #1:

1) Figures are missing axis labels and alphabetic labels (e.g. A, B which are used in the figure caption) throughout the manuscript and supplementary materials. For example, in Figure 1 and Figure 4. Please add these to ensure the figures are clear.  
  
2) Figure 4 caption refers to a legend which isn’t present. Please add and check other figures.  
  
3) Figure S12 refers to ‘institutions’, this terminology is not used anywhere else in the manuscript. The distinction between institution and hospital is not clear to me, please clarify.

Reviewer #2: No

Reviewer #3: No

5. If applicable, are the interpretation of results and study conclusions supported by the data?  
  
Please provide suggestions (if needed) to the author(s) on how to improve, tone down, or expand the study interpretations/conclusions. Please number each suggestion so that the author(s) can more easily respond.

Reviewer #1: Mark as appropriate with an X:  
Yes [X] No [] N/A []  
Provide further comments here:

Reviewer #2: Mark as appropriate with an X:  
Yes [X] No [] N/A []  
Provide further comments here:

Reviewer #3: Mark as appropriate with an X:  
Yes [X] No [] N/A []  
Provide further comments here:  
  
The estimation procedure in itself is correct, but I have problems with several assumptions of the model.

6. Have the authors clearly emphasized the strengths of their study/theory/methods/argument?  
  
Please provide suggestions to the author(s) on how to better emphasize the strengths of their study. Please number each suggestion so that the author(s) can more easily respond.

Reviewer #1: Yes, the importance of modelling nosocomial transmission, particularly of bacterial pathogens is clear, as are the study findings. A few minor comments regarding the inference methodology:  
  
1) The abstract would benefit from stating that an ensemble adjustment Kalman filter is used.  
  
2) A rationale is needed for using the ensemble adjustment Kalman filter method over a method more often used for inference of transmission rate parameters, such as ABC/SMC/MCMC.

Reviewer #2: Yes.

Reviewer #3: Yes

7. Have the authors clearly stated the limitations of their study/theory/methods/argument?  
  
Please list the limitations that the author(s) need to add or emphasize. Please number each limitation so that author(s) can more easily respond.

Reviewer #1: 1) The study period coinciding with peak COVID-19 cases in the city is mentioned as a limitation. Further discussion on how this might affect the study results would be beneficial. Did you consider fitting multiple transmission rates for different time periods, perhaps when different levels of restrictions were in places?  
  
2) A more detailed discussion is needed on the types of infectious pressure that the ABM model is not capturing, particularly the lack of hospital staff in the model. These are briefly mentioned in Understanding nosocomial transmission.  
  
Ideally when modelling nosocomial transmission an ABM would either explicitly model HCWs and hospital staff as agents in the model, or implicitly include them within the force of infection. Moreover, infection dynamics could be disentangled further if other sources were considered, such as: staff infectious pressure, spatial infectious pressure from other wards (patients/staff) that may be connected, and a background infection pressure. I assume staff data was unavailable for this study?  
  
  
3) It is unclear to me why the force of infection for each ward only considers the infection/colonization status of patients in the same ward. Did you consider adding a ‘background’ transmission rate parameter which scaled with the number of infections in the hospital at that time?  
  
  
4) A daily contact network was used rather than a finer-scale (e.g. hourly) to capture patient movements, was this for computational benefit? It is not clear to me if a patient was admitted into A&E and subsequently to an inpatient ward on the same day, whether they would only appear in the inpatient ward of the contact network? If so, the model may not be capturing transmission in these staging areas.

Reviewer #2: Yes, but some of these limitations could be addressed by a global sensitivity analysis as suggested.

Reviewer #3: Yes

8. Does the manuscript structure, flow or writing need improving (e.g., the addition of subheadings, shortening of text, reorganization of sections, or moving details from one section to another)?  
  
Please provide suggestions to the author(s) on how to improve the manuscript structure and flow. Please number each suggestion so that author(s) can more easily respond.

Reviewer #1: The manuscript flows well, just one very minor point:  
1) The final three sentences of the introduction seem unnecessary and should be covered in the abstract/results and discussion.

Reviewer #2: Yes.  
The layout of the manuscript is excellent and the subheadings are very clear and make the manuscript very easy to navigate. However the writing in this manuscript feels rushed and is confusing in places. The manuscript would benefit from a thorough readthrough and edit by someone less familiar with it's content.

Reviewer #3: No

9. Could the manuscript benefit from language editing?

Reviewer #1: No

Reviewer #2: Yes

Reviewer #3: No

More information and support   
  
FAQ: How do I revise my submission in Editorial Manager?  
https://service.elsevier.com/app/answers/detail/a\_id/28463/supporthub/publishing/  
  
FAQ: How can I reset a forgotten password?   
https://service.elsevier.com/app/answers/detail/a\_id/28452/supporthub/publishing/   
For further assistance, please visit our customer service site: https://service.elsevier.com/app/home/supporthub/publishing/  
Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live chat and email  
  
At Elsevier, we want to help all our authors to stay safe when publishing. Please be aware of fraudulent messages requesting money in return for the publication of your paper. If you are publishing open access with Elsevier, bear in mind that we will never request payment before the paper has been accepted. We have prepared some guidelines (https://www.elsevier.com/connect/authors-update/seven-top-tips-on-stopping-apc-scams ) that you may find helpful, including a short video on Identifying fake acceptance letters (https://www.youtube.com/watch?v=o5l8thD9XtE ). Please remember that you can contact Elsevier s Researcher Support team (https://service.elsevier.com/app/home/supporthub/publishing/) at any time if you have questions about your manuscript, and you can log into Editorial Manager to check the status of your manuscript (https://service.elsevier.com/app/answers/detail/a\_id/29155/c/10530/supporthub/publishing/kw/status/).  
  
#AU\_EPIDEMICS#  
  
To ensure this email reaches the intended recipient, please do not delete the above code