PEER REVIEW HISTORY

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

ARTICLE DETAILS

TITLE (PROVISIONAL)	Studying the time trend of methicillin-resistant Staphylococcus aureus (MRSA) in Norway by use of non-stationary Gamma-Poisson distributions.
AUTHORS	Moxnes, John; Moen, Aina; Leegaard, Truls

VERSION 1 - REVIEW

REVIEWER	Karim Khader University of Utah Salt Lake City, Utah USA
REVIEW RETURNED	05-May-2015

GENERAL COMMENTS	I think this is a nice contribution to the study of MRSA trends, and highlights many of the challenges associated with controlling the spread of MDRO. 1. In the abstract under Objectives (page 2, line 11), recommend changing the wording in the major question from "Is the number of MRSA in Norway..." to perhaps "Is the number of MRSA isolates in Norway..." 2. In the abstract, under Setting (page 2, lines 26-27) I do not understand the description of dataset II, in particular, how the two time periods are related. Are all MRSA isolates collected in Health Region East from 2002 - 2008 and 2002 - 2011 the same as all isolates collected in Health Region East from 2002 - 2011? What is the reason for listing two overlapping time periods? 3. I would recommend identifying a reference for each of the first two sentences in the Introduction (page 4, lines 8-10): "... colonizing about a third of the world's population." and "S. aureus has the ability to quickly develop resistance to antimicrobial agents." 4. I would recommend replacing "MRSA" on page 5, lines 12, 14, 22 with "MRSA infections (or cases or counts)" or something to that effect. 5. It would be good to briefly describe why "The advantage is that the proportion is difficult to collect correctly." and "MRSA count is more trustworthy." 6. As in the abstract, I am still unclear what the purpose of describing dataset II as MRSA isolates collected from 2002 - 2008 and 2002 - 2011. 7. Page 6 line 11: I would replace the term "MRSA" with "MRSA

	isolates" 8. Page 6 line 45: What is the implication of the assumptions that MRSA isolates are Poisson distributed together with the assumption of independent MRSA tests? Is this an approximation? 9. I would recommend removing equation (2.2) from the manuscript and simply use words to define the mean and variance (e.g. The mean and variance are given by \mu and \sigma^2 respectively). 10. I do not follow the section on "Model Estimation" (page 7, lines 41 - 51). What is the message in this section? What is the contribution of equation (2.3) to the message? 11. I would recommend defining the notation used in equation (2.3). Alternatively, perhaps this step can be described with words with regards to how it relates to "Model Estimation". 12. What are x_i^sim and \tilde{x_i}^sim, how are they defined and calculated (page 8, line 3 and equation (2.4))? 13. I would recommend removing the links to the wikipedia pages for Overdispersion and Gamma distribution on page 8. 14. I recommend removing equation (2.5) from page 8 and leaving it in the appendix for the readers reference. 15. The authors suggest that both domestic and import cases are increasing (page 11, lines 11-12). Is it possible to theorize to what degree the import and domestic cases feed off of each other? Also, are there implications of either imported or domestic cases on the 5-year projections? 16. Although the rates took specific functional forms (exponential, linear), yet MRSA case counts depend on the underlying population, it would be interesting to know whether the Poisson rate function could be defined explicitly in terms of the change in population/demographics. Is that a possibility? 17. It is known that the false negative probability of MRSA cultures is non-negligible. Are there information about the MRSA culture sensitivity and specificity that were used to test these isolates, and what do the authors think the impact of missed MRSA cases due to false negatives might have on their study?

VERSION 1 – AUTHOR RESPONSE

Reviewer Name Karim Khader
Institution and Country University of Utah
Salt Lake City, Utah
USA
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

I think this is a nice contribution to the study of MRSA trends, and highlights many of the challenges associated with controlling the spread of MDRO.

AC: Thank you we agree.

1. In the abstract under Objectives (page 2, line 11), recommend changing the wording in the major question from "Is the number of MRSA in Norway..." to perhaps "Is the number of MRSA isolates in Norway..."

AC: Corrected.

2. In the abstract, under Setting (page 2, lines 26-27) I do not understand the description of dataset II, in particular, how the two time periods are related. Are all MRSA isolates collected in Health Region East from 2002 - 2008 and 2002 - 2011 the same as all isolates collected in Health Region East from 2002 - 2011? What is the reason for listing two overlapping time periods?

AC: We are sorry for not having seen this error in our writing. We have collected data from one time period and have inserted the following sentence: ‘The second dataset (dataset II) consists of all MRSA isolates collected in Health Region East from 2002-2011.’

3. I would recommend identifying a reference for each of the first two sentences in the Introduction (page 4, lines 8-10): "... colonizing about a third of the world's population." and "S. aureus has the ability to quickly develop resistance to antimicrobial agents."

AC: We agree and have inserted the following references:

Wertheim HF, Melles DC, Vos MC, et al. The role of nasal carriage in Staphylococcus aureus infections. Lancet Infect Dis 2005;5(12):751-62.
and
Lowy FD. Antimicrobial resistance: the example of Staphylococcus aureus. J Clin Invest 2003;111(9):1265-73.

4. I would recommend replacing "MRSA" on page 5, lines 12, 14, 22 with "MRSA infections (or cases or counts)" or something to that effect.

AC: Corrected.

5. It would be good to briefly describe why "The advantage is that the proportion is difficult to collect correctly." and "MRSA count is more trustworthy."

AC: We have inserted the following sentence:

We have collected all MRSA findings in the South-Eastern part of the country over long time periods. This makes the MRSA count in the time period studied reliable. MSSA is not mandatory notifiable in Norway making it more difficult to extract the MRSA proportion.

6. As in the abstract, I am still unclear what the purpose of describing dataset II as MRSA isolates collected from 2002 - 2008 and 2002 - 2011.

AC: Once again we are sorry for not having seen these wrongly written sentences.
The main article, data section, now states ‘The second dataset (dataset II) consists of all MRSA

isolates collected in Health Region East from 2002-2011. Health Region East consists of Oslo County and four neighboring counties and is the most populated area of Norway; it includes many large and small cities and rural areas, and covers approximately 36% of the Norwegian population. MRSA isolates from Oslo County are included in both datasets, but due to the use of two different databases for extraction of data, SWISSLAB (Swisslab, GmbH, Berlin, Germany) for dataset I and MICLIS (Miclis AS, Lillehammer, Norway) for dataset II, and due to other factors outlined and discussed in our previous paper, 9 the two datasets for Oslo county could not be combined.'

7. Page 6 line 11: I would replace the term "MRSA" with "MRSA isolates"

AC: Corrected

8. Page 6 line 45: What is the implication of the assumptions that MRSA isolates are Poisson distributed together with the assumption of independent MRSA tests? Is this an approximation?

AC: We now write:

The Poisson process is a natural starting point for count data analysis. The non-stationary Poisson process is a next step when modeling non stationary processes though it is often inadequate due to overdispersion.21

9. I would recommend removing equation (2.2) from the manuscript and simply use words to define the mean and variance (e.g. The mean and variance are given by \mu and \sigma^2 respectively).

AC: We do not think it is a good idea to remove this equation from the manuscript. The equality of mu and lamda is important here. However, we have simplified the equation by removing the definition of mean and variance.

10. I do not follow the section on "Model Estimation" (page 7, lines 41 - 51). What is the message in this section? What is the contribution of equation (2.3) to the message?

AC: We have inserted the following sentence: In this model estimation section we establish the parameters of the stochastic model that is used for forecast. The realization of the stochastic process in equation (3) is used to compare with the given data and for forecast.

11. I would recommend defining the notation used in equation (2.3). Alternatively, perhaps this step can be described with words with regards to how it relates to "Model Estimation".

AC: We have inserted the following sentence for definition: Equation (3) gives the simulated number of MRSA for each month where we numerically draw a random number from a Poisson distribution with intensity parameter lamda that varies with time.

12. What are x_i^sim and \tilde{x_i}^sim, how are they defined and calculated (page 8, line 3 and equation (2.4))?

AC: We have inserted the following text after equation (4) [Previously (2.4)]:

Here x_i is the measured number of MRSA counts, and x_i^sim is the simulated number of MRSA counts.

13. I would recommend removing the links to the wikipedia pages for Overdispersion and Gamma distribution on page 8.

AC: We have removed the links to Overdispersion and Gamma distribution.

14. I recommend removing equation (2.5) from page 8 and leaving it in the appendix for the readers reference.

AC: We prefer to keep the equations since they are used in equation (6) [previously (2.6)]. We have moved both on one line to minimize space.

15. The authors suggest that both domestic and import cases are increasing (page 11, lines 11-12). Is it possible to theorize to what degree the import and domestic cases feed off of each other? Also, are there implications of either imported or domestic cases on the 5-year projections?

AC: To this date, no studies in Norway have shown whether the increase in domestic cases are a result of imported cases, but as we already state, this has been shown in Sweden, and a similar situation would be logical in Norway too, as both MRSA levels, societies and healthcare systems are similar. It in unknown whether the import and domestic cases feed off of each other, and we would rather not speculate.

16. Although the rates took specific functional forms (exponential, linear), yet MRSA case counts depend on the underlying population, it would be interesting to know whether the Poisson rate function could be defined explicitly in terms of the change in population/demographics. Is that a possibility?

AC: We have in the conclusion inserted the following sentence:

The number of MRSA counts may depend on the population size if the intensity of the Poisson process increases with population number. However, we have seen no sign of such a relationship.

17. It is known that the false negative probability of MRSA cultures is non-negligible. Are there information about the MRSA culture sensitivity and specificity that were used to test these isolates, and what do the authors think the impact of missed MRSA cases due to false negatives might have on their study?

AC: In the conclusion section we have added the following:

False negative probability is negligible since the sensitivity of both culture and molecular detection methods are very good. We believe that not testing for MRSA is by far a much larger problem.30

Reviewer Name Thomas Neyens
Institution and Country I-Biostat (Hasselt University, Belgium)

The authors present an interesting study. Their written English is good and the text follows a clear structure.

AC: Thank you, we agree.

I especially value their clear overview of study limitations.
AC: Thank you.

I think this work can be valuable, but a number of issues have to be solved. I have a few major and some minor comments:

AC: OK

Major:

- The biggest problem here is that nowhere in the statistical analysis, a correlation between time points is taken into account. The gamma Poisson model is indeed capable of modelling extra variability, but it does not take into account the fact that much of the heterogenity may be caused by the correlation within the data. Please investigate the use of possible alternatives that consult this issue or clearly explain why these techniques have not been considered.

AC: We have inserted the following sentence:

The overdispersion in our data may be due to failure of the assumption of independence or correlation of events. Indeed, particular assumptions of observed heterogeneity due to dependence and correlation lead to the gamma Poisson distribution.22 In general we believe that the overdispersion is due to individuals acting as a group (i.e. epidemic outbreaks). However, individual responses to covariates such as dates may also be an explanation.

- LSF is used, while ML is not used. While this is rightfully mentioned as a limitation, I would like the authors to give clear arguments as to why this method was not considered.

AC: We add the following sentence in the conclusion:

A diagnostic plot of the empirical fit of the variance using LSF on squared residuals suggested that the overdispersion in the data was significant. Other diagnostic fits may more directly underscore the maximum likelihood principle.

- Can possible inter-hospital variability not be taken into account by modeling counts per hospital? In that way, you shift towards a longitudinal setting where you can introduce a random effect to capture this type of correlation. Please investigate to possibilities of using such models.

AC: We added the following sentence:

Inter-hospital variability may be modeled in future research where different random effects may be accounted for separately. This will increase insight on how demographics influence MRSA development.

- Maybe I am missing something, but the explanation on Study II (p5 bottom en p6 top) is not clear. As I understand it correctly, these data are separated in 2 parts, according to database changes in 2008. But is this a correct interpretation? You're talking about data from 2002-2008 and 2002-2011. But there is an overlap between those two periods. What am I missing? Can you please write it down more clearly.

AC: We are sorry for not having seen these wrongly written sentences. Both in the abstract, p. 2, line numbers 25-28, and in the main article, data section, p. 5 line numbers 52-59 and p.6 line numbers 3-5, the faults have been corrected.

The abstract now states: ‘The second dataset (dataset II) consists of all MRSA isolates collected in Health Region East from 2002-2011’,
The main article, data section, now states ‘The second dataset (dataset II) consists of all MRSA isolates collected in Health Region East from 2002-2011. Health Region East consists of Oslo County

and four neighboring counties and is the most populated area of Norway; it includes many large and small cities and rural areas, and covers approximately 36% of the Norwegian population. MRSA isolates from Oslo County are included in both datasets, but due to the use of two different databases for extraction of data, SWISSLAB (Swisslab, GmbH, Berlin, Germany) for dataset I and MICLIS (Miclis AS, Lillehammer, Norway) for dataset II, and due to other factors outlined and discussed in our previous paper, 9 the two datasets for Oslo county could not be combined.'

Minor:

- Formula (2.1): Isn't it better to use lambda_i instead of lambda (as you say it is time dependent)?

AC: We have inserted lamda_i to show the time dependency.

- I would suggest that you use 3.1 Model Identification, 3.2 Model Estimation and 3.3 Diagnostic Checks on Model Adequacy and Overdispersion instead of the formulation without numbers.

AC: We have removed the numbering of the section headings to be in agreement with the journals setting.

- Please carefully look at spelling possible typo's, for example on p.5 (line 17), there should be a space between '7' and 'the'.

AC: Corrected.

VERSION 2 - REVIEW

REVIEWER	Karim Khader Research Assistant Professor Division of Epidemiology, University of Utah United States of America
REVIEW RETURNED	15-Jul-2015

GENERAL COMMENTS	Line 174: ", where t_0 the is number of..." probably intended to be ", where t_0 is the number of..." Line 188: "of observing X_i = x_i" has both lower and upper case x with a subscript i, yet the following equation does not have a subscript i for lower case x. Line 198: A definition of "def" is provided and seems to suggest that "def" appears in equation (2), but it does not seem to be the case that "def" appears before its definition. Equation (4): I think that it would be helpful to provide some text describing what is meant with the notation LSF(Diff^{Exp}) and LSF(Diff^{Sim}).

VERSION 2 – AUTHOR RESPONSE

Reviewer Name Karim Khader
Institution and Country Research Assistant Professor
Division of Epidemiology, University of Utah
United States of America
Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below
Line 174: ", where t_0 the is number of..." probably intended to be ", where t_0 is the number of..."
AC: Corrected.

Line 188: "of observing X_i = x_i" has both lower and upper case x with a subscript i, yet the following equation does not have a subscript i for lower case x.
AC: Corrected to X_i = x

Line 198: A definition of "def" is provided and seems to suggest that "def" appears in equation (2), but it does not seem to be the case that "def" appears before its definition.
AC: The sentence ‘where “def” refers to definition.’ has been deleted

Equation (4): I think that it would be helpful to provide some text describing what is meant with the notation LSF(Diff^{Exp}) and LSF(Diff^{Sim}).
AC: We have inserted the following text on page 8, lines 233-236: LSF(Diff^{Exp} is the LSF to the squared difference between the MRSA data and the estimated intensity based on the MRSA data. LSF(Diff^{Sim}) is the LSF to the squared difference between the simulated data and the estimated intensity based on the simulated data.