Reviewer’s report

Title: Chlamydia trachomatis infection rates among a cohort of mobile soldiers stationed at Fort Bragg, North Carolina, 2005-2010.

Version: 2 Date: 11 June 2013

Reviewer: Sereina Herzog

Reviewer’s report:

This manuscript is interesting and useful to the scientific community and is therefore suitable for publication in BMC Public Health. Having said this however, I did find the discussion difficult to follow. There are some major points to be clarified and also several minor points. Especially, references are missing in several places of the discussion. I am hopeful that you will consider including some suggestions.

Major Compulsory Revisions

Methods

1. The description or explanations of the categories within the characteristics are missing. For example, how are the categories “single” and “non-single” in “marital status” defined or what is the reason to combine ages 26-67 into one category whereas the “17-20” and “21-25” are separated?

2. Does the proportional hazards assumption hold for the different exposures? In this context it could be helpful to see Kaplan-Meier curves, e.g. pay grade, race/ethnicity (provided as supplementary material).

Results

3. It should be highlighted how the multivariable model differ between men and women (and discuss the differences in the discussion).

4. The adjusted HR/multivariable model for recurrent Ct infection differs from incident Ct, and even more between men and women – describe the differences shortly and discuss it in discussion.

Discussion

5. I did find the discussion difficult to follow. The first paragraph starts with a summary and comparison, ends with a statement about generalizability of the result, and is followed by a paragraph with discussion about presented results. For example, I would expect to read the paragraphs about the crude estimates before the characteristics are discussed and compared with other studies.

6. The discussion what are plausible reasons that the multivariable model differs between men and women is missing for the incident and for the recurrent Ct infection and the disparities between incident and recurrent Ct model.

7. First paragraph, the statement “The crude incidence rate of Ct infection …” is difficult to place in context without a value or a reference. However, the
information comes in a later paragraph but the reference and a relation (twice as high) should be provided with the statement. The reference(s) should also be provided for the statement “With exceptions of descriptive analyses based on national ...”.

8. First paragraph, in the manuscript no data or references are presented which support the statement (last sentence of the paragraph) that the military population represents the U.S. population well. And in addition no information/references are presented how representative Fort Bragg is for the U.S. military population. I would suggest phrasing the sentence more as a potential strength of the study in the context of strengths and limitations of the study.

9. Fifth paragraph, please add reference for the first sentence about the Armed Forces Health Surveillance Center or otherwise why not report the crude rate for all active personnel during the study period?

10. Fifth paragraph, what is the reference for the female Army study by Gaydos – and is it Ct prevalence or incidence? And in the following sentences, are the 7.99 and the 3.91 incidence rates per 1,000 person-years and do they refer to the female Army study by Gaydos?

Minor Essential Revisions

General

11. In several places the word “protected” is used to describe an observed reduced hazard ratio, e.g. in the abstract “having prior deployments protected against both incident and recurrent infection”. I would suggest rewording these phrases as protection reads to describe a cause rather than to indicate direction. For example, “prior deployments were associated with a lower hazard of both ...”.

12. It would be helpful if always the same wording would be used to describe and discuss the results about the characteristics and the corresponding categories:

- In the abstract it is written “unmarried” whereas in the tables marital status is divided into “single” and “non-single”.
- In results (2nd para of ‘incident infection’), “who deployed for a military operation” is written in the text whereas the description used in the tables is “Deployment during study”. The same for the statement “who were otherwise away from the garrison setting for an extended period” whereas this characteristic in the tables is described by “Breaky in duty at Ft. Bragg”.
- In results (2nd para of ‘incident infection’), in text “lower rank” – in tables “pay grade” and again “non-married status” whereas in the tables the characteristic is grouped into “single” and “non-single”.
- In results (2nd para of ‘recurrent infection’), “history of any prior overseas deployment” in the text – in tables “history of prior deployments.”
- In discussion (1st para), “black race” – in combination with method section this should be “non-white” and again “not being married” – in tables “non-single”.

Abstract

15. Chlamydia trachomatis should be written in italic.

Background

16. The shortcut for Chlamydia trachomatis should be introduced within the background.

Methods

17. Which program was used for the analyses?

18. Second paragraph of ‘Study location and Population’, please state in the method section why Ct diagnosis less than 30 days apart were omitted. The explanation in the discussion section is too late.

19. First paragraph of ‘Statistical analysis’, I assume that the first part about the Defense Medical Surveillance System belongs rather to the ‘Data sources’ part of the method section.

20. First paragraph of ‘Statistical analysis’, I would suggest to use the terminology “univariate” and “multivariate” or “univariable” and “multivariable”. The unadjusted hazard ratio analyses with one characteristic involve two variables (CT incident and the characteristic) but there is only one independent variable and therefore the term “bivariate” should be replaced. In a strict statistical sense, multivariate analyses means the study of how several outcome variables vary together but the these analyses are often referred to as multivariate methods (see BR Kirkwood and JAC Sterne “Essential Medical Statistics”).

21. The method section ‘Statistical analysis’ misses the information that the analysis was done stratified by gender.

22. End of first and start of second paragraph of ‘Statistical analysis’, it is a bit confusing how censoring events and follow-up time were defined for the incident Ct and recurrent Ct. I would suggest provide the definitions first for the incident Ct followed by the information for the recurrent Ct.

Results

23. First paragraph, depending on the shape of the distribution it could be necessary to provide the median instead of the mean for the follow-up time. It would be interesting to see the distribution of the follow-up time stratified by gender (maybe in a supplementary file).

24. First paragraph of ‘Incident infection’, is the crude Ct incident really 22.1? Using the information provided in table 1: 2198/101149.9 results in 21.7 per 1,000 person-years.

25. First paragraph of ‘Incident infection’, last sentence comparing men and women, are these crude incidences? If so please include the word “crude”.

13. Please be consistent with the use of “to” or “-“ to indicate the range of CIs. (e.g. 1st para of ‘incident infection’ in the results section).


16. The shortcut for Chlamydia trachomatis should be introduced within the background.

17. Which program was used for the analyses?

18. Second paragraph of ‘Study location and Population’, please state in the method section why Ct diagnosis less than 30 days apart were omitted. The explanation in the discussion section is too late.

19. First paragraph of ‘Statistical analysis’, I assume that the first part about the Defense Medical Surveillance System belongs rather to the ‘Data sources’ part of the method section.

20. First paragraph of ‘Statistical analysis’, I would suggest to use the terminology “univariate” and “multivariate” or “univariable” and “multivariable”. The unadjusted hazard ratio analyses with one characteristic involve two variables (CT incident and the characteristic) but there is only one independent variable and therefore the term “bivariate” should be replaced. In a strict statistical sense, multivariate analyses means the study of how several outcome variables vary together but the these analyses are often referred to as multivariate methods (see BR Kirkwood and JAC Sterne “Essential Medical Statistics”).

21. The method section ‘Statistical analysis’ misses the information that the analysis was done stratified by gender.

22. End of first and start of second paragraph of ‘Statistical analysis’, it is a bit confusing how censoring events and follow-up time were defined for the incident Ct and recurrent Ct. I would suggest provide the definitions first for the incident Ct followed by the information for the recurrent Ct.

23. First paragraph, depending on the shape of the distribution it could be necessary to provide the median instead of the mean for the follow-up time. It would be interesting to see the distribution of the follow-up time stratified by gender (maybe in a supplementary file).

24. First paragraph of ‘Incident infection’, is the crude Ct incident really 22.1? Using the information provided in table 1: 2198/101149.9 results in 21.7 per 1,000 person-years.

25. First paragraph of ‘Incident infection’, last sentence comparing men and women, are these crude incidences? If so please include the word “crude”.
26. Second paragraph of ‘Incident infection’, I would suggest to change the sentence “were all associated with Ct infection among female …” to “were all associated with higher Ct infection rates among female …”

27. First paragraph of ‘Recurrent infection’, is the crude recurrent infection rate 118.0? Using the information provided in table 2: 223/2013.31 results in 110.7 per 1,000 person-years. Please add 95% CI for this estimate.

Discussion

28. First paragraph, history of prior deployments was found to result in lower rates in the univariable model for men and women but only for women in the multivariable model. Therefore the first sentence would only be true for the univariable analysis which should be clarified.

29. Fifth paragraph, please give results round off to one decimal place in this paragraph.

30. Seventh paragraph, the statement “… only found to hold in women after adjustment for all covariates.” is not correct, the multivariable model for the women does not contain all covariates listed in the table.

Table 1

31. As the analyses are preformed stratified by gender I would suggest splitting the general columns “Total” and “PY” by gender. The forth column (PY) misses the total person-years.

32. For readability of the table it would be helpful to represent values bigger then thousand using thousands separators.

33. The fifth column (Infected) shows also the percentages, please indicate this in the column heading or write the percent sign next to the values.

34. The race/ethnicity characteristic is shown in the univariable model with four categories whereas in the multivariable model it is included only differentiating “white” and “non-white” as described in the method section. Please include the information for the univariable hazard ratio for this characteristic as it is used for the multivariable model. The detailed information for the total, PY, and infected for the “non-white” could be still provided under the category “non-white” using an indentation for the subcategories “Hispanic”, “Black”, and “Other”. The characteristic in the multivariable model for men includes three hazard ratios. This is contradicting to what was described in the method section.

35. For the first category in the characteristic “Months in service at entry” the time unit is not needed to be repeated.

36. The percentages for the age categories in women do not sum up to 100%.

37. The characteristic “Months at service at entry” misses hazard ratios in the multivariable model in the men and in women.

38. The characteristic “Primary occupation specialty at entry” misses in the multivariable model for women the 1.00 for the reference group.

39. The multivariable model for women misses for the characteristic age the 1.00 for the reference group and for the men there is no HR for the category “21-25”.
40. See also comments 31 to 35.
41. The multivariable model for men misses for the characteristic ‘Months in service at entry’ the 1.00 for the reference group.

Discretionary Revisions
42. As in comment 20, I would recommend renaming the columns about unadjusted and adjusted hazard ratio, e.g. a column name for both models like “Hazard ratio (95% CI)” and then subheading for each model “Univariable” and “Multivariable”.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.