Author's response to reviews

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

Authors:

Shilpa Hakre (shakre@hivresearch.org)
Robert J Oyler (robert.j.oyler@us.army.mil)
Kenneth A Ferrell (kenneth.ferrrell@us.army.mil)
Fang Li (fli@hivresearch.org)
Nelson L Michael (nmichael@hivresearch.org)
Paul T Scott (pscott@hivresearch.org)
Bruno P Petruccelli (bpetruccelli@hivresearch.org)

Version: 4 Date: 6 December 2013

Author's response to reviews: see over
We thank the referees for their time and are grateful to both referees for the thoroughness of their review and their constructive comments. We agreed with a majority of their recommendations and revised the manuscript accordingly which has led to a significantly improved manuscript. Our response to each of the referees’ comments is indicated below in blue font.

Referee #1

Reviewer’s report:

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?

Rationale for categorization and descriptions for categories have been added to the Statistical Analysis.

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).

We used the log rank test of significance as well as Kaplan-Meier survival curves to check whether the proportional hazards assumption was met for >13 variables each for males and females in univariate analysis for both incident and recurrent Ct analyses for a total of >52 variables prior to entry into multivariate Cox models. These exploratory analyses that were performed have now been described in the Statistical Analysis section.

Results

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

We did not add discussion points to address sex-related differences. Rather the more logical flow in the Discussion makes clearer the interpretation of such differences based on what is known regarding the differences both in the expected proportion of asymptomatic infection and in the likelihood of screening. As a related point to illustrate this, sex (gender) itself was chosen as a basis for dividing models, rather than as a covariate in one overall model, precisely because Ct infection is (in a practical sense) a different entity in women compared to men, particularly as such infections are identified in military populations.

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.

We did not add discussion points to address differences between the incidence- and recurrence-based models. Rather, the subdividing of tables now more clearly reminds the reader that the denominators for recurrent infection are a subset of the overall, person-time figures representing everyone assigned to Fort
Bragg during the period of study; and this, in itself, accounts for actual differences between the models. As for how the risk factor analysis differs between incident and recurrent cases, the more logical flow of the Discussion should improve reader understanding of the latter as a higher risk subgroup.

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.

The Discussion paragraphs have been re-ordered for a more logical and conventional flow, as results are addressed first with respect to findings overall and then in relation to covariates.

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. [See the responses to numbers 3 and 4 above.] We did not add discussion points to address sex-related differences and differences between the incidence- and recurrence-based models. Rather, the subdividing of tables and the more logical flow in the Discussion now makes clearer the interpretation of such differences.

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 …”.

In rearranging the Discussion paragraphs, the “crude incidence” statement is both proximate to its appropriate context and the reference that had already accompanied that context (but had previously not been iterated for the isolated, more general statement in beginning of the previously submitted version of the Discussion). Two references have also been added to exemplify studies with large N’s based on national and regional data (Datta & Xu).

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.

This statement has now been moved to the end of the Discussion as a counterpoint to immediately preceding remarks on study limitations. The intent was not to claim demographic representativeness of the US population; but rather to emphasize the geographic diversity (“every region contributes…”), and thus suggest that any pre-existing host factors are not unique to Fort Bragg or even North Carolina. This makes the analysis similar to college-based studies, but contrasts with community- or city-based studies. Unfortunately, a more
detailed discussion would inordinately lengthen the paper.

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?

A reference has now been added.

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?

A reference for each of these two statements is now numbered in the appropriate locations. (Unbeknown to us the reference manager software program in use at the time of manuscript submission was corrupted. Therefore all references accepted from track changes in final formatting were lost).

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 …”.

Throughout the document, the word “protected” has now been replaced (e.g., by an appropriate use of “associated” or “association”).

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”.

Consistent terms are now used (e.g., preferring “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 “Breaks in duty at Ft. Bragg”.

More consistent language is now used, and other breaks in duty are distinguished from deployments. (Where wording/usage is more appropriate for the full text, some liberty is retained in using the past tense verb “deployed” rather than the noun form “deployment” that is more appropriate for table row labels.)

- 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”.

Consistent language is now used.

- In results (2nd para of ‘recurrent infection’), “history of any prior overseas deployment” in the text – in tables “history of prior deployments”.

Where wording/usage is more appropriate for the full text, some liberty is retained in using adjectival modification to ensure the reader understands that military deployments in most cases (and certainly in this study) refer to overseas movements. Such modification or detail is then not necessary for the table.

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”.

Consistent terms are now used (e.g., preferring “single”). Where a less stratified value, such as “non-white”, is the intended grouping for analysis, the appropriate term is retained.

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).
These have now been made consistent.

These have now been made consistent.

Abstract

15. Chlamydia trachomatis should be written in italic.
The originally intended italics have now been restored for occurrences of full genus-and-species names.

Background

16. The shortcut for Chlamydia trachomatis should be introduced within the background.
The abbreviation is now iterated (reintroduced) in the Background section so that the full text is independent of the Abstract with respect to internal definitions.

Methods

17. Which program was used for the analyses?
We used SAS and have now provided this information in the Statistical Analysis section.

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.
The rationale for excluding Ct diagnoses less than 30 days apart has been added to a newly created Case Definition section and a reference has been provided .[Johnson RE, et al.: Screening tests to detect Chlamydia trachomatis and Neisseria gonorrhoeae infections--2002. MMWR Recomm Rep 2002, 51(RR-15):1-38].
The 2010 STD Treatment Guidelines published by the U.S. Centers for Disease Control and Prevention indicate that nucleic acid amplifications diagnostic tests can detect non-viable Ct DNA up to 3 weeks after treatment. We chose a 30-day cutoff in defining recurrent infection and excluded any diagnoses within the defined interval to avoid inclusion of potentially persistent infections.
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.

This entire paragraph has been moved to the end of the Data Sources 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 these analyses are often referred to as multivariate methods (see BR Kirkwood and JAC Sterne “Essential Medical Statistics”).

We have adjusted the statistical term to the style preferred by the reviewer and now use “univariable” both in the Statistical Analysis section and Table titles.

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

We revised the section by indicating that all analyses were conducted with stratification for 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.

We redefined censoring and clarified methods of calculating follow-up time for incident or recurrent Ct cohorts.

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).

We revised descriptive statistics provided in the text and replaced the mean with median times. Moreover, to provide readers more information about the distribution, we added the interquartile ranges for all medians described in the text. Additionally, we added medians (and interquartile ranges) in describing follow-up time for females and males.

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.

Thank you for pointing out this discrepancy. We have corrected this to 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”. Since the comparison was of crude incidence rates, the word “crude” has been added.

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 …”

This sentence has now been changed to indicate association with higher infection rates.

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.

We rectified the discrepancy in calculation and omission of CI Results.

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.

In the new location of these statements, a parenthetical clause was added to point out the non-sex-specific finding in univariate analysis (“though in univariate analysis such history was associated with a lower incidence rate in both women and men”).

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

Where appropriate, numbers have now been rounded to one decimal place.

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.

This is now clarified as “after adjustment for significant covariates”.

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.

We agree with the reviewer. The Total and PY for females and males are provided separately and the Total years of person-time have been added to each column.

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

We chose not to use separators in compliance with BMC Public Health’s instructions to authors regarding numbers in tables (“Commas should not be used to indicate numerical values.” <http://www.biomedcentral.com/bmopublichealth/authors/instructions/researcharticle#preparing-tables>).

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.

We inserted headers to the columns to indicate whether the numbers are n’s or percentages.

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.

We appreciate the suggestions for clarification and have inserted the group level “Non-white” above the sublevel groups of Black, Hispanic, and Other. The total n, person-time, frequencies, percentages, and hazard ratios have been added to the tables illustrating univariate analyses.

We clarify in the Statistical Analysis section in Methods that for multivariate analysis race categories were collapsed based on similarity of risk levels.

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

The redundancy has been remedied. The word “months” has been deleted from the first level in the characteristic “Months in service at entry”.

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

We have fixed the percentages so the rounded numbers add up to 100%.

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

The first two levels of the characteristic “Months in service at entry” were collapsed for statistical power in the recurrent Ct multivariate Cox model. As a result, since the table lacks gridlines, it appears as if the hazard ratios are missing. We have created two more tables presenting just the multivariate analyses. The new presentation of results make clearer that values are not missing.

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

Thank you for noticing the omission. The 1.00 has now been added to indicate 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”.

Again, thanks for uncovering omission. We have now added 1.00 to indicate the referent age group in females. The 17-20 and 21-25 age groups for males were collapsed into one group for the multivariate model therefore it appears as if values are missing. Per response in 37 above, the addition of multivariate-only tables should make clearer that values are not missing.
Table 2

40. See also comments 31 to 35.

We have made similar changes as indicated in responses to comments 31-35 to Tables 2-4 in the revised manuscript. In response to the second reviewer’s suggestion, we split the incident and recurrent infection tables into 4 tables which present multivariate analyses separately from univariate analyses.

41. The multivariable model for men misses for the characteristic ‘Months in service at entry’ the 1.00 for the reference group.

The last group in the characteristic “Months in service at entry” has now been indicated as the referent level.

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”.

As indicated in response 41, we present the univariate and multivariate analyses in separate tables. The title for each table indicates the type of analysis i.e. univariate or multivariate.
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: 9 September 2013

Reviewer: Sharon Kuhlmann

Reviewer's report:

Comments are given by page number, paragraph (par xx) and line within paragraph (line xx); line number is given as +xx meaning number of lines from first line of the paragraph, or as –xx, meaning lines from last line of paragraph.

Major Compulsory Revisions

1. Page 5, par 2, line 6: Where only one recurrent infection per person considered, i.e. the first one that occurred at least 30 days after the incident infection?

   We clarify the definition of recurrent Ct by adding that it is the first recurrent diagnosis 30 days or more after the incident Ct diagnosis that was considered.

2. Page 7, par 1, line -4: If test of cure was being conducted, why was it necessary to remove Ct infections within 30 days of another infection? Does this mean that that data was not used?

   The 2010 STD Treatment Guidelines published by the U.S. Centers for Disease Control and Prevention indicates that nucleic acid amplifications diagnostic tests can detect non-viable Ct DNA up to 3 weeks after treatment. We chose a 30-day cutoff in defining recurrent infection to exclude diagnoses possibly due to persistence of infection.

3. Page 7, par 2, line -5: How were non-white race-ethnicity categories combined and why? It is not clear in the text nor in Table 1. Also age seems to have been combined in Table 1/Men but it is not mentioned or motivated in the Statistical analysis section.

   We added an explanation in Methods that covariate groups were collapsed for multivariate analyses if risk levels were contiguous.

4. Page 7, par 2, line -5: It is not clear how recurrent infections were modeled, specifically what the outcome was since one person might have had several recurrent infections. Was only the first recurrent infection included? If so, how does this imply limitations to the study? Alternatively, were all infections included? If so, how was time calculated for the second, third, … infections? And how was the repeated measurements aspect taken into account?

   We chose to examine only the time to the diagnosis of the first recurrent infection during the study period due to the sparsity of subsequent recurrence observed in the cohort. (Only 31 of 2,104 soldiers met the definition for a second recurrent infection and three for a third.) Studying the time to the first recurrent infection and factors associated with this event would best inform prevention efforts at Fort Bragg as the first recurrence occurred with greatest
frequency and contributed most to the burden of Ct infections at the installation. We clarified in Methods that only the first diagnosis of recurrence was studied. Additionally, we added the following to Results to describe the burden of subsequent recurrent infections:

“During follow-up, 31 (13.9%) of the 223 with an initial recurrent infection had a second recurrent diagnosis of Ct and 3 (1.3%) had a third.”

5. Page 7, par 2, line -3: The term “censoring events” is currently confusing. The event being studied, here Ct infection, is usually called event in survival analysis, while the conditions for being excluded from the cohort are called censoring (as in “censoring was defined as ….”).

We clarified the definition of censoring in the Statistical Analysis section and make clear the outcome was not a censoring event.

6. Page 8, par 1, line -2: What type of test was used to compare? This should be mentioned in the Statistical analysis section.

We expanded the Statistical Analysis section and inserted that Pearson chi-square and Kruskal-Wallis tests of significance were used to compare categorical and continuous variables, respectively.

7. Page 8, par 2, line 2: It is not clear in this paragraph if only incident infections are included or all Ct infections. For example the sentence “The final number of infections excluded 72 that occurred within 30 days of a prior Ct diagnosis.” seems to indicate that incident and recurrent infections are included in the total, similarly the following sentence “The number of individuals found to have had at least one Ct infection during the study period was 2,198, or 3.3% of the cohort.”

We revised the beginning of the paragraph as follows:

“Among 67,425 soldiers assigned to Fort Bragg during January 1, 2005 to June 14, 2010, 2,198 individuals (3.3%) were reported to have had at least one Ct infection. The total number of both initial and recurrent (n=223) infections, excluding 72 that occurred within 30 days of a prior Ct diagnosis, was 2,493.”

8. Page 8, par 2, line 4: If the paragraph only refers to incident infections, and 2493 persons reported an incident infection with 72 lost to follow up, then 2493-72=2421 persons should be included in the study. However, the next sentence states “The number of individuals found to have had at least one Ct infection during the study period was 2,198, or 3.3% of the cohort.” It is not clear where does 2198 come from.

The revised sentence in response to comment 7, which indicates the number of recurrent infections, addresses the discrepancy in the total.

9. Page 8, par 2, line 7: “50% received their Ct diagnosis within 6 months of return to Fort Bragg (…)”: do these numbers in the sentence refer only to incident infections or any kind of Ct infection?

We revised the sentence to clarify that the 50% belonged to the incident infection group only.

10. Page 8, par 2, line -4: Does the crude incidence rate refer to incident infections only?

Yes, the crude incidence rate is for incident Ct only and we revised the sentence to indicate this.
11. Page 8, par 2, line -3: Is the number of person-years of follow up mentioned when comparing incidence rates in women and men the same as the denominator used when calculating the crude incidence-rate in the previous sentence?

For males and females, we used their respective person-years of follow-up time in calculating crude incidence rates. Reviewer 1 suggested gender-specific person-times replace the total person-years presented. These person-years are now provided in the univariate/unadjusted tables which you suggested we create in your Discretionary Revisions comment 4 below.

12. Page 9, par 1, line 6: What type of test was used to compare? This should be mentioned in the Statistical analysis section.

As mentioned in our response to comment 6, we added the types of statistical test used, in this case Pearson chi-square, to the Statistical Analysis section.

13. Page 10, par 1, line 4: there is not mention to the incidence rates previously calculated for active duty personal nor a reference is given either.

We revised the first paragraph and added the previously reported rates and a reference.

14. Where the assumptions of proportional hazard checked?

We used Kaplan-Meier survival curves and the log rank test of significance to assess whether the assumptions of proportional hazard were met. The Statistical Analysis section has been revised to indicate as such.

15. It is not clear if covariates were updated (e.g. age, marital status, deployment) when analyzing recurrent infections.

We clarified this in the text as follows:

"We used a Cox proportional hazards model to assess the association of covariates at study entry (assignment at Fort Bragg) with time to incident Ct infection; covariates were remeasured prior to entry into the recurrent Ct cohort (i.e. at time of incident Ct infection) to assess their association with time to first recurrence."

Minor Essential Revisions

1. Page 4, par 2, line 6: "Ct" abbreviation has not been defined previously in the article.

The definition for Ct has been added to the Background section.

2. Page 5, par 1, line 2: The size of the active duty population should be added to provide context.

The size of the active duty population at Fort Bragg was added to the Study Location and Population section and a reference was provided.

3. Page 7, par 1, line 1: What type of screening is carried out at Fort Brags? Is it target to any special group?

We provide more detail about the Defense Medical Surveillance System that it is a system (relational database) that captures all medical and service-related information about an individual throughout their service in the military and not only when they are at a single installation like Fort Bragg.

4. Page 8, par 1, line 4: when reporting mean age, it would be preferable to report median age as usually age has a skewed distribution (which is apparent
here when comparing range and mean).

In agreement with your suggestion, we provided the median and interquartile range throughout the Results section.

5. Page 8, par 1, line 6: Also when reporting follow-up time, I would like to see the median time.

The median has been included to replace the mean and this sentence has been moved to the Incident Infection section to which it refers.

6. The software used for statistical analysis is not mentioned.

We used SAS for data management and analysis and have revised the Statistical Analysis section to indicate as such.

Discretionary Revisions

1. Page 5, par 2, line 3: For clarity of presentation, I suggest that the text starting “Ct infection were ascertained … reliable protocol” should be included under the heading of “Case definition” or “Outcome” as it is not “Study location and population”.

We added a Case Definition section and all pertinent text have been moved to this section.

2. Page 6, par 2, line 2: The sentence on the protocol review should be moved to the end as it breaks the flow as it is now.

We agree and have moved the sentence on protocol review as recommended.

3. Data sources section could be shortened by removing information that is not relevant for the study, such as serological testing of other diseases.

The information on serological testing and follow-up for other diseases has been removed.

4. Table 1 and 2: I suggest breaking up the table in two: the first should include Total and Unadjusted estimates for women and men, and a second table with include adjusted results for women and men.

Thank you for the suggestion. We agree and separate presentation of univariate and multivariate analyses from two tables into a total of four tables.

5. Page 7, par 3, line +1: The first sentence of the paragraph belongs to the previous paragraph as it refers to censoring definition.

We moved the sentence in question to the previous paragraph as suggested.

6. Page 5, par 4, line -2: The sentence “The population at ...” describes more clearly the objective of the study than what is currently stated in the background section. You might consider moving this sentence earlier for the benefit of the reader.

We agree and accordingly moved the sentence to the first paragraph in Study Location and Population.