Reviewer’s report

Title: Difference in overall and age-specific prevalence of human papillomavirus infection in Italy: evidence from NTCC trial

Version: 2  Date: 13 August 2012

Reviewer: Christine Pierce Campbell

Reviewer’s report:

The authors use data from a randomized clinical trial of cervical cancer screening in Italy to calculate the overall (age-standardized) and age-specific prevalence of high-risk HPV across 9 geographical areas of northern and central Italy. In addition to estimating HPV prevalence, the authors use joinpoint analyses to describe changing patterns of HPV infection across the lifespan. Overall, the manuscript does not seem to provide a useful contribution to scientific knowledge regarding the HPV epidemiology. Perhaps this study fills a gap in HPV knowledge in Italy, however, the authors do not make this statement. Furthermore, I am concerned with the use of age-specific prevalence measures in joinpoint analyses; joinpoint analyses are typically used to assess time trends in age-standardized incidence or mortality rates, or occasionally, time trends in prevalence. I believe the authors are attempting to demonstrate a non-linear relationship between age and HPV prevalence for each geographical area; however, I do not know if joinpoint analyses are the most appropriate method for this aim. The authors may want to consult a statistician if they are unsure of the applicability of their method.

- Major Compulsory Revisions

1. Methods, fourth paragraph. To assess heterogeneity across study centers, you used meta-analysis methods. Why did you decide to use this particular method instead of a grouped randomized control trial method, since all centers were within the same study (NTCC)?

2. Methods, fifth paragraph. Confirm that the use of joinpoint analysis is appropriate for the study aims.

- Minor Essential Revisions

In general: For consistency, use one term to describe the HPV types examined (high-risk, oncogenic, or carcinogenic). For each instance of HPV in the results and discussion, preface the word with high-risk/oncogenic/carcinogenic (e.g., high-risk HPV prevalence) since the authors did not examine low-risk HPV types.

1. Title. Include the term “high-risk” HPV in the title, since the analyses included only high-risk HPV types.

2. Abstract, methods. Briefly describe the intervention arm of the NTCC study. The words “changes in prevalence trends by age” is confusing and implies a time component. Perhaps use “differences in the age-distribution of HPV infection.”
Mention there are 9 centers.

3. Abstract, results. “No significant joinpoints were observed” provides little to no information to a reader unfamiliar with joinpoints. “Age slopes did not differ significantly” from what…across centers?

4. Abstract, conclusions. A summary of the findings would be helpful.

5. Background, second paragraph. The first two sentences review literature on HPV incidence and persistence. Perhaps a review of HPV prevalence literature would be more useful. The third sentence is unclear; why can (can’t) we infer the natural history of HPV from age distribution? If you are referring to geographical variation, perhaps this should be explicitly stated. Also, mention why it is important to consider geographical variation in age-specific prevalence of HPV.

6. Background, first paragraph. The first sentence states that previous studies were unable to study cohort effects; this implies that the present study examines cohort effects, which it does not. The aim/purpose of this study is unclear. The last paragraph states what was done, however, it should be written as an aim instead. The use of the words “age shape of HPV prevalence” is confusing. Perhaps say “age-distribution of HPV prevalence.”

7. Methods, fourth paragraph. First sentence should add “…was considered in determining HPV prevalence.” Second sentence could read as follows, “The overall age-standardized HPV prevalence for each center was estimated by direct standardization using the standard female Italian population for the year 2004, truncated from 25-60 years.” Third sentence could read, “The heterogeneity in age-standardized HPV prevalence across centers was tested using the [NAME] method proposed by …” Name and describe this method, along with why you chose to use meta-analysis methods to analyze heterogeneity across centers within the same study (NTCC). Also, I could not find this data in the results section.

8. Methods, fifth paragraph. Which methods were used to calculate the age-distribution of HPV prevalence? Again, “the shape of age-specific prevalence” is confusing; use “The age-distribution of HPV prevalence.” A brief overview of joinpoint analysis should be provided, as not all readers will be familiar with this method. Briefly mention the purpose of joinpoint analysis and what can be achieved through its use, along with a definition/interpretation of a statistically significant joinpoint. The following source provides a clear description of joinpoint methodology:

http://www.portal.state.pa.us/portal/server.pt?open=18&objID=901769&mode=2

The R package is “ljr” not “lrj”. Also, the statistical software (R) itself should be cited (see below). Furthermore, you should state the name of the statistical test used within the “ljr” package, if available.


With which methodology did you categorize age into groups? State that the level of statistical significance was set at p<0.05.
9. Results, first paragraph. Mention that the 47,369 women were in the NTCC study. Clarify that the “HPV group” is the HPV testing arm. Perhaps say “…46,900 remained and were included…” Are there any other participant characteristics that you can report (e.g., median age)?

10. Results, second paragraph. Present data from Figure 2 prior to Table 1, as Figure 2 provides the most general data (HPV prevalence by center) compared with Table 1 (HPV prevalence by center and age). A statistical test for heterogeneity (?) is mentioned, but this was not described in the methods. Also, this test is not provided as part of Figure 2. Perhaps modify the second sentence to say “Age-standardized high-risk HPV prevalence varied significantly (p<0.001) across centers with … All remaining centers had values between 8.2% and 9.4%.” I could not find the numbers 8.0% and 9.5% listed in the manuscript. Table 1 provides some interesting information that should be mentioned in the results text. For example, there was an overall high-risk HPV prevalence of 8.4%, prevalence varied geographically (range: 5.7%-9.7%), and for all centers the prevalence of high-risk HPV was highest among women 25-29 years of age and lowest among women 50-60 years of age.

11. Results, third paragraph. Suggestion: “Joinpoint regression analyses suggest that no significant joinpoints occurred in age-specific high-risk HPV prevalence in 6 centers…” Which statistical test was used to test for heterogeneity in intercept? Slope? It appears as though the intercept is for women age 25, but you mention this is for overall prevalence; please clarify.

12. Results, fourth paragraph. Explain the “fitted regression”; is this a line of best fit for the data? Why are you fitting a line to the data? What does it represent? This was not mentioned in the methods section. It may be helpful to present the data for Turin, then Ravenna, and finally Padua. Modify the order of centers in Figure 4 accordingly. For Revenna, you mention “low values” and “high values”, but do not define what is considered low or high.

13. Discussion, first paragraph. Instead of saying that you observed different prevalences, perhaps say “We observed geographical variation in age-standardized and age-specific prevalence of high-risk HPV infection…almost twice that of Trento (a mainly rural mountain area), the lowest prevalence.” Are there any data to support the claimed variation in sexual behavior across Italy? If so, please cite these studies.

14. Discussion, second paragraph. In the first sentence, “…non-negligible differences were found…in the shape of age specific prevalence.” Again, perhaps it would be better to say “…differences were found in the age-distribution of high-risk HPV prevalence across areas.” What does non-negligible mean? Are you referring to a statistical test? Instead of age-pattern, I would say age-distribution. I would be careful of using “Persistence of high prevalence” as the term “persistence” has a special meaning in the HPV literature. How are you defining “less young women”? In the fifth sentence, how did you come to these conclusions? The connection is not immediately apparent. Perhaps replace the word husbands with partners, if you did not specifically ask for their marital status or sexual orientation.
15. Discussion, third paragraph. In the second sentence, you mention changes in sexual behavior, but do not describe these “changes”. For what changes are you referring? Also, please elaborate on the smoking habits.

16. Discussion, fourth paragraph. Elaborate on these thoughts. They sound incredibly interesting but are not well-developed, coherent sentences.

17. Discussion, fifth paragraph. Shouldn’t population mixing decrease geographical variation? Perhaps you are referring to a rural-to-urban migration (or vise versa) that may explain an increase in the variation of HPV prevalence. If so, please clarify.

18. Discussion, sixth paragraph. Suggestion: “The NTCC study population is large and is considered to be representative of Italian women between the ages of 25 and 60 years who participate in cervical cancer screening…” In the third sentence you mention no significant effects of education, marriage, or place of birth; were any significant effects found in other variables? Fourth sentence: I suggest removing the word “especially”, and explaining why you believe selection bias to be unlikely. In the weaknesses, it is important to state that you have no data from southern Italy. The final weakness listed (i.e., power) should be mentioned at the beginning of your weaknesses.

19. Conclusions. I did not see data on differences in age-specific HPV prevalence across centers. Also, I did not see data for birth cohort effects. The final two sentences are interesting and important, and should have been mentioned in both the background and in the discussion sections. These statements require some explanation, and would be best addressed in the background/discussion.

20. Table 1. Title: Prevalence of high-risk cervical HPV by center and age. Column heading, “HPV+/Tested women (%).” Whole numbers should have a decimal place (e.g., 1.0%).

21. Figure 2. Title. The age-standardized prevalence of high-risk cervical HPV, by recruitment center. Provide a p-value for test of heterogeneity.

22. Figure 3. Title. Model-based age-specific prevalence of high-risk cervical HPV for centers with linear age trends. Y-axis should be, “High-risk HPV prevalence (%).”

23. Figure 4. Title. [Model-based?] age-specific prevalence and fitted regression line of high-risk cervical HPV for centers with non-linear age trends. Y-axes should be, “High-risk HPV prevalence (%).” Re-order panels based on appearance in text.

- Discretionary Revisions

1. Abstract, background. Clarify that the decreasing prevalence of HPV with increasing age is among women. This is not true among men.

2. Background, first paragraph. “HPV infection varies between populations.” I would use the term “across” rather than between.

3. Methods, first paragraph. The first sentence could add (Figure 1) to the very
end, and delete the words “shows the location of…” List the 9 centers in the same order that their population size is described. Adriatic Sea is without an “o”.

4. Methods, second paragraph. In the first sentence, specify that it is the NTCC study.

5. Methods, third paragraph. Instead of “In the HPV group”, I would recommend saying “Among women in the HPV testing arm”, which is a bit more specific. What is phase 1 and 2, and is it important to mention here? Instead of “targeting any of the oncogenic HPV types” I would say “was used to identify any of the high-risk HPV types … and was considered positive at 1 relative light unit.” There is no need to add (RLU) since the term is not used again in the manuscript. “High reproducibility of HPV testing between laboratories…” should say “across laboratories…”

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:

I declare that I have no competing interests