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)	Prevalence of human papillomavirus in Wenzhou, China: a cross-sectional study of 127,938 outpatient women
AUTHORS	Zhang, Mengqi; Chen, Gang; Dai, Xuchao; Wu, Zhigang; Huang, Hong; Zheng, Yuanyuan

VERSION 1 – REVIEW

REVIEWER	Brouwer, Andrew University of Michigan, Epidemiology
REVIEW RETURNED	18-Aug-2022

GENERAL COMMENTS	Summary: This study assessed prevalence of HPV in cervical samples collected in a city in China from 2011 to 2020. The study is a largely straightforward assessment of prevalence by age, year, and genotype. The strengths of the study include the large sample and longitudinal data. The primary weakness of the study is in the statistical details. Overall, I think that the study has worthwhile information to add to the literature but has some flaws that need to be addressed. Major comment: The paper has multiple statistical weaknesses. These include the following: • The number of genotypes that were tested for changed from 21 to 27, but the authors don't account for that fact when calculating the overall prevalence (i.e., the prevalence would artificially increase just because they're detecting more genotypes). The two periods either need to be analyzed separately, or prevalence for the latter period needs to be restricted to the smaller set of genotypes. • The authors show age- and time- trends in prevalence, but don't include an adjusted prevalence accounting for the fact that the distribution of ages is not constant year-to-year. A column of adjusted prevalence should be included in Tables 1 and 2. • The authors say that they make a Bonferroni adjustment when comparing prevalence across ages and across years, but they do not apply it to the Joinpoint analyses, where I think it is especially needed. • It doesn't seem sensible to me to time trend analysis on having 2, 3, 4, or 5 genotypes, specifically. The real problem is that prevalence of, e.g., 2 genotypes specifically is tied to the prevalence of having one genotype, and the real question is whether the prevalence of each genotypes is independent of the others or not. The analysis plan here needs more thought. Minor comments: -English language editing is needed.

	-Abstract. The abstract needs to state how many and which HPV genotypes were assayed, otherwise we can't understand the prevalence. -Abstract. "Fluctuated" is not the right word. I think you want "varied." Fluctuate implies a rise and fall that I don't think you intend, since you later describe specific time trends. -Ln 70 & Abstract. The authors repeat in a couple of places that prevention and control of HPV at a local level rely on an understanding of the prevalence and genotype distribution. I don't think that this is the case. Regardless of the prevalence and genotype distribution, our control measures are vaccination and promotion of safe sex. Understanding genotype distributions are broadly useful to understand the ecology of the disease and whether vaccine updates might be needed, but I don't think it drives local control efforts. -Ln 86. HPV is not a disease of low-resource settings specifically, so it's not clear that high living standards are in contrast with a moderate HPV prevalence. -Patient involvement. The study design is unclear at this point. Are the authors simply accessing the results of HPV tests starting in 2011. Or, are the authors accessing banked samples and testing them? If so, did the patients consent to banking the samples? Please clarify the study design here. -Ln 107. The authors indicate that only patients from "better-off families" had HPV testing. Doesn't this induce substantial selection bias? Some discussion of this point is warranted. -Ln 161. Rather than describing the trend as "flat," which it is not, perhaps say "the trend was not statistically significant." -Ln 216. The bimodal pattern has been seen in numerous studies around the world (US, Latin America, perhaps others), and it would be good to cite some of this broader literature. -Supplement 3. Why are the percentages given as a total of the whole population in that year? I would like to see them given by age group. It's not clear what "n" is referring to in this table. Population totals? Numbers of cases?

REVIEWER	Iamaroon, Anak Chiang Mai University
REVIEW RETURNED	29-Sep-2022

GENERAL COMMENTS	The article is informative, interesting and well written. The below are my comments. 1. Did the patients have HPV vaccination status since this may affect the results of the study? 2. Risk behaviors like multiple sexual partners were not included in the study. Is this information available? The authors should discuss the risk and the possibility of an increased rate of HPV infection in younger patients. 3. The article would be more valuable if the data of the socioeconomic status of the patients were included in the study.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1:

Dr. Andrew Brouwer, University of Michigan

Comments to the Author:

Summary: This study assessed prevalence of HPV in cervical samples collected in a city in China from 2011 to 2020. The study is a largely straightforward assessment of prevalence by age, year, and genotype. The strengths of the study include the large sample and longitudinal data. The primary weakness of the study is in the statistical details. Overall, I think that the study has worthwhile information to add to the literature but has some flaws that need to be addressed.

Major comment:

The paper has multiple statistical weaknesses. These include the following:

- The number of genotypes that were tested for changed from 21 to 27, but the authors don't account for that fact when calculating the overall prevalence (i.e., the prevalence would artificially increase just because they're detecting more genotypes). The two periods either need to be analyzed separately, or prevalence for the latter period needs to be restricted to the smaller set of genotypes.

Response: This comment is greatly appreciated, and we greatly agree with you. We analyzed the two periods together, with the prevalence of the latter period restricted to a smaller set of genotypes. We corrected all the relevant data and recalculated them. Although it took a lot of effort, we appreciate your suggestions considering the rigor of epidemiological statistical analysis. Please see line 146-148.

Line 146-148:

"This study was a trend analysis over a long period of time, so only 21 genotypes were also selected for statistical analysis in the latter time period in order to avoid bias in the overall prevalence."

- The authors show age- and time- trends in prevalence, but don't include an adjusted prevalence accounting for the fact that the distribution of ages is not constant year-to-year. A column of adjusted prevalence should be included in Tables 1 and 2.

Response: We appreciate your comments. We share your view that age-standardized prevalence rates would be more meaningful. Age standardized prevalence rates would eliminate differences in the age composition of the population and allow for better comparison with different regions. However, we have little data for the standard age structure population in Wenzhou to calculate the standardized prevalence rate.

- The authors say that they make a Bonferroni adjustment when comparing prevalence across ages and across years, but they do not apply it to the Joinpoint analyses, where I think it is especially needed.

Response: Thank you for the heads up. We performed Bonferroni adjustment for all time trend analyses. we added the relevant content in the revised manuscript. Please see line 207-209 and 222-224.

Line 203-205:

"We compared data from different years for each age group and found no significant difference between years after Bonferroni correction."

Line 218-210:

"We compared the data of each genotype in different years, and there was no significant difference between each year after Bonferroni correction."

- It doesn't seem sensible to me to time trend analysis on having 2, 3, 4, or 5 genotypes, specifically. The real problem is that prevalence of, e.g., 2 genotypes specifically is tied to the prevalence of having one genotype, and the real question is whether the prevalence of each genotypes is independent of the others or not. The analysis plan here needs more thought.

Response: Thank you for this comment. Firstly, we agree with you. We removed the relevant content of the temporal trend analysis of multiple genotype infections. Secondly, we added Table 4 for a descriptive analysis of the overall prevalence of multiple genotypes. We refer to the presentation of the results of a similar study that has been published (*Li et al.,2020*). We hope that some of the revisions will improve the quality of the manuscript. Thank you again for your comments. Finally, the current epidemiological studies on HPV genotypes, mainly in the description of age, time, genotype distribution. Further research is related to cervical cancer. On the question of whether the prevalence of each genotype is independent of other genotypes, we think this is a good idea. Our current research focuses on time trend analysis, but we think your idea is very novel and is a good entry point for future research. For additions and modifications, please see line189-197.

Li HP, Li PQ, Huang LY, et al. Prevalence characteristics of cervical human papillomavirus (HPV) infection in the Zhoupu District, Shanghai City, China. Virol J. 2020,17(1):84.

Line 189-197:

"3.4 Prevalence of single and mixed HPV infection

Single HPV genotype infection was found to be the most common pattern (74.28%, 16,910/22,766). The multiple genotype infection rate was 25.72% (5,856/22,766). Single-type infection (74.28%) was more common than multiple-type infection (25.72%) (Table 4). Regardless of the year, single HPV infection was the most common, and the prevalence was much higher than the sum of multiple infections (Supplement 3)."

Table 4: Prevalence of single and multiple HPV infection

Genotype of HPV infection	Number of cases	Prevalence (%)	Percentage of positive samples (%)
Single[a]	16910	13.2	74.28
Double	4455	3.48	19.57
Triple	1028	0.80	4.52
Quadruple	278	0.22	1.22
Quintet	65	0.05	0.29
Sextuple and more	30	0.02	0.13

[a]The prevalence of single HPV infection was higher than that of multiple HPV infection. $p<0.001$.

Minor comments:

-English language editing is needed.

Response: Thanks for your comment. This manuscript has been grammatically corrected, but we are willing to further improve by means of manuscript editing services for English if necessary.

-Abstract. The abstract needs to state how many and which HPV genotypes were assayed, otherwise we can't understand the prevalence.

Response: Thanks for the reminder. We have supplemented the relevant content in the abstract. Please see line 40-42.

Line 40-42:

"The 21 HPV genotypes included in this study were HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 6, 11, 42, 43, 44 and 81."

-Abstract. "Fluctuated" is not the right word. I think you want "varied." Fluctuate implies a rise and fall that I don't think you intend, since you later describe specific time trends.

Response: Thanks for your correction. We have replaced "fluctuated" to "varied ". Please see line 48-49.

Line 48-49:

"The annual prevalence varied between 12.2-28.8%, with a significant decrease in the years 2011-2018 and a flat trend in 2018-2020."

-Ln 70 & Abstract. The authors repeat in a couple of places that prevention and control of HPV at a local level rely on an understanding of the prevalence and genotype distribution. I don't think that this is the case. Regardless of the prevalence and genotype distribution, our control measures are vaccination and promotion of safe sex. Understanding genotype distributions are broadly useful to understand the ecology of the disease and whether vaccine updates might be needed, but I don't think it drives local control efforts.

Response: We agree with you. We have made changes to the inappropriate content. Please see line 33-34, 56-57,79-82 and100-102.

Line 33-34:

"Understanding the prevalence and genotype distribution of human papillomavirus (HPV) is of great significance for HPV vaccination and updating."

Line 56-57:

"In addition, we need to promote HPV vaccination and increase the publicity of safe sex."

Line 79-82:

"For a region or city, it is helpful to understand the ecology of the disease and whether the HPV vaccine needs to be updated by mastering the trend of its incidence and the distribution characteristics of viral genotypes."

Line 100-102:

“We hope that the results of this study will help to understand the ecology of HPV-infected diseases and provide a scientific basis for the development and promotion of vaccines.”

-Ln 86. HPV is not a disease of low-resource settings specifically, so it’s not clear that high living standards are in contrast with a moderate HPV prevalence.

Response: Thanks for your correction. The original expression is not very rigorous, we deleted the relevant content of the economic level. It has been modified in the revised manuscript. Please see line 96-97.

Line 96-97:

“A previous study showed that HPV infection in the region was at a moderate level in China [18].”

-Patient involvement. The study design is unclear at this point. Are the authors simply accessing the results of HPV tests starting in 2011. Or, are the authors accessing banked samples and testing them? If so, did the patients consent to banking the samples? Please clarify the study design here.

Response: Thanks for your comment. We obtained the HPV test results of outpatients in Wenzhou People 's Hospital from 2011 to 2020. We added research design related content in the revised manuscript. Please see line 118-123.

Line 118-123:

“We obtained the HPV test results of outpatients in Wenzhou People 's Hospital from 2011 to 2020. This study is based on established data, does not involve interventions with patients, and exemptions from informed patient consent have been requested from the ethics committee. The data are anonymous, and the requirement for informed consent was therefore waived. Patients or the public were not involved in the design, conduct, reporting, or dissemination plans of our research.”

-Ln 107. The authors indicate that only patients from “better-off families” had HPV testing. Doesn’t this induce substantial selection bias? Some discussion of this point is warranted.

Response: Thanks for the heads up. We added relevant content in the discussion. Please see line 319-323.

Line 315-319:

“Some limitations are remaining in this study. First, the sample includes only women who had been to a gynecological clinic, which might lead to higher results. Second, HPV testing is only a test recommendation based on the needs of the patient's condition and economic level, and the principle of patient voluntariness may cause selection bias.”

-Ln 161. Rather than describing the trend as “flat,” which it is not, perhaps say “the trend was not statistically significant.”

Response: Thank you again for the correction. It has been modified in the revised manuscript. Please see line 170-174.

Line 170-174:

“By means of Joinpoint time trend analysis, the annual percentage change (APC) was -10.9 (95% CI: -13.7 to -8.0, P=0.001) from 2011 to 2018 and 21.8 (95% CI: -3.8 to 54.2, P=0.09) from 2018 to 2020,

indicating a significant decrease in the years 2011-2018 and the trend was not statistically significant in 2018-2020."

-Ln 216. The bimodal pattern has been seen in numerous studies around the world (US, Latin America, perhaps others), and it would be good to cite some of this broader literature.

Response: Thanks for your comment. We added references to related research in other countries. Please see line 241-242.

Line 237-238:

"This bimodal pattern is consistent with other studies at home and abroad [20,22-25]."

[23] Lorenzon L, Terrenato I, Dona MG, et al. Prevalence of HPV infection among clinically healthy Italian males and genotype concordance between stable sexual partners. J Clin Virol. 2014; 60:2649.

[24] Bruni L, Diaz M, Castellsagué X, et al. Cervical human papillomavirus prevalence in 5 continents : meta-analysis of 1 million women with normal cytological findings. J Infect Dis, 2010, 202(12):1789-1799.

[25]de Sanjosé S, Diaz M, Castellsagué X, et al. Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology : a meta- analysis. Lancet Infect Dis, 2007, 7(7):453-459.

-Supplement 3. Why are the percentages given as a total of the whole population in that year? I would like to see them given by age group. It's not clear what "n" is referring to in this table. Population totals? Numbers of cases?

Response: We appreciate your comments. We changed the presentation of the results of Supplement 3 according to your opinions. The percentage denominator here is changed to the total number of HPV positive cases per year. The corresponding data changes were modified in the results and discussions. Due to the need to modify the original manuscript, Supplement 3 was changed to Supplement 4 in the revised manuscript. Please see Supplement 4.

Reviewer: 2

Dr. Anak Iamaroon, Chiang Mai University

Comments to the Author:

The article is informative, interesting and well written. The below are my comments.

1. Did the patients have HPV vaccination status since this may affect the results of the study?

Response: We appreciate your comments. We also believe that whether patients are vaccinated with HPV vaccine has a great impact on HPV test results. However, we only obtained HPV test results from 2011 to 2020 from outpatients in Wenzhou People 's Hospital. Meanwhile, little is known about the patient 's personal information. We don't know if they're vaccinated.

2. Risk behaviors like multiple sexual partners were not included in the study. Is this information available? The authors should discuss the risk and the possibility of an increased rate of HPV infection in younger patients.

Response: Thanks for your comment. Due to the limitations of the original data, we cannot obtain information such as the patient 's behavior habits. However, we believe that the study of the impact of risk behavior habits on HPV infection is valuable, and we will conduct related research in the next step. We discussed the risk and possibility of increased HPV infection in young patients in the revised manuscript. Please see line 250-254.

Line 246-250:

"Young women have a higher prevalence of HPV, which may be related to inappropriate sexual behavior [28]. The detection of HPV in women has been found to begin consistently with a peak immediately after the onset of sexual relations, usually from 15 years of age, mostly transient infection, which can be cleared within 1 or 2 years [29]."

[28] Zhao FH, Tiggelaar SM, Hu SY, et al. A multi-center survey of age of sexual debut and sexual behavior in Chinese women: suggestions for optimal age of human papillomavirus vaccination in China. Cancer Epidemiol. 2012; 36: 384–90.

[29] Bruni L, Diaz M, Castellsagué X, et al. Cervical human papillomavirus prevalence in 5 continents: meta-analysis of 1 million women with normal cytological findings. J Infect Dis 2010; 202:1789–99.

3. The article would be more valuable if the data of the socioeconomic status of the patients were included in the study.

Response: Thanks for the heads up. Unfortunately, we do not have data on the socioeconomic status of patients, but we are in favor of your idea. In our next study, we will take into account.

Editor's comments:

- Please revise the title of your manuscript to include the research question, study design and setting. This is the preferred format of the journal.

Response: Thanks for your comment. We have revised the title as requested.

- Please ensure that your abstract is formatted according to our Instructions for Authors: http://bmjopen.bmj.com/pages/authors/#research

Response: Thanks for the heads up. We ensure that the abstract format conforms to the Instructions for Authors.

- Please add a section entitled 'Strengths and limitations of this study' (immediately after the abstract). This section should contain up to five short bullet points, no longer than one sentence each, that relate specifically to the methods. The novelty, aims, results or expected impact of the study should not be summarised here.

Response: Thanks for your comment. We have added a section entitled 'Strengths and limitations of this study' (immediately after the abstract).

- The PPI statement should not contain details of participant recruitment; this information should be included elsewhere in your Methods section. Please see our blog for further information regarding PPI. http://blogs.bmj.com/bmjopen/2018/03/23/new-requirements-for-patient-and-public-involvement-statements-in-bmj-open/

Response: Thanks for your comment. We have put the details of participant recruitment elsewhere in the method section.

- All items from the checklist should be included in your manuscript. Please do not leave blanks and indicate any items that do not apply to your study design as 'Not Applicable'.

Response: Thanks for your comment. We have checked all the items in the checklist and made changes in inappropriate places.

- Please complete a thorough proofread of the text and correct any spelling and grammar errors that you identify.

Response: Thanks for your comment. This manuscript has been grammatically corrected, but we are willing to further improve by means of manuscript editing services for English if necessary.

VERSION 2 – REVIEW

REVIEWER	Brouwer, Andrew University of Michigan, Epidemiology
REVIEW RETURNED	26-Oct-2022

GENERAL COMMENTS	The authors edits have largely addressed my comments. A few minor concerns remain: 1. The captions and column headings for many of the tables (Table 3, S3, S4) are insufficient to understand what the numbers presented mean, e.g., not clear what "strains" means in Table 3 and the difference between proportion and prevalence; not clear if n is sample size of number of positives in S3 and S4. 3. I recommend strengths and limitations be addressed in the discussion rather than in the conclusion section. 4. English language and style editing is still recommended, including for the new title.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Dr. Andrew Brouwer, University of Michigan

Comments to the Author:

The authors edits have largely addressed my comments. A few minor concerns remain: 1. The captions and column headings for many of the tables (Table 3, S3, S4) are insufficient to understand what the numbers presented mean, e.g., not clear what "strains" means in Table 3 and the difference between proportion and prevalence; not clear if n is sample size of number of positives in S3 and S4.

Response: We appreciate your comments. We changed the title of Table 3. We replaced column heading " Number of strains (Strains)" with "Number of positive samples (cases)", " Proportion (%)" with " Proportion among HPV-positive samples (%)", and " Prevalence (%)" with " Prevalence in total samples (%)". Please see line 186-187. In Supplement 3, "n" refers to the total number of samples

per year. In Supplement 4, "n" refers to the number of positive samples. Below each table we have added explanations.

Line 186-187:

Table 3 Prevalence and distribution of each HPV gene subtype.

HPV subtypes Number of

positive samples

(cases) Proportion among HPV-positive samples (%) Prevalence in total samples (%)

HR-HPV 24670 79.4 19.3

16 3788 12.2 3.0

18 1509 4.9 1.1

31 776 2.5 0.6

33 1189 3.8 0.9

35 483 1.6 0.4

39 1552 5.0 1.2

45 376 1.2 0.3

51 1115 3.6 0.9

52 4592 14.8 3.6

53 2458 7.9 1.9

56 961 3.1 0.8

58 3152 10.1 2.5

59 949 3.1 0.7

66 1000 3.2 0.8

68 770 2.5 0.6

LR-HPV 6420 20.6 5.0

6 1552 5.0 1.2

11 1344 4.3 1.1

42 278 0.9 0.2

43 777 2.5 0.6

44 746 2.4 0.6

81 1723 5.5 1.3

3. I recommend strengths and limitations be addressed in the discussion rather than in the conclusion section.

Response: Thanks for your comment. We have put the strengths and limitations related content into the discussion section. Please see line 303-314.

Line 303-314:

“In summary, this study assessed HPV infection characteristics and the time trend among gynecological outpatients in Wenzhou from 2011 to 2020 with a large sample size and a long time series. There were some limitations. First, the sample only included women who had been to a gynecological clinic, which might lead to higher results. We obtained only HPV test results and could not ascertain whether the patient was vaccinated for HPV. HPV vaccination has a significant impact on HPV test results. There was no information about risky behavior. Second, recommendations for HPV are based on the patient's needs and economic level, and the voluntary nature might cause selection bias. Third, we do not have data on socioeconomic status; we will consider this in future studies. Finally, the lack of cervical cytology or histological results makes it impossible to correlate HPV infection and genotype distribution with different cervical abnormalities.”

4. English language and style editing is still recommended, including for the new title.

Response: Thanks for the heads up. Our manuscript has been revised by a professional editor from a language retouching agency. Some grammatical and manuscript editing errors have been corrected. The title has also been changed.