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)	Epidemiology, etiology and seasonality of infectious diarrhea in adult outpatients through active surveillance in Shanghai, China, 2012-2016: a cross-sectional study
AUTHORS	Gong, Xiaohuan; Wu, Huan-yu; Li, Jian; Xiao, Wen-jia; Zhang, Xi; Chen, Min; Teng, Zheng; Pan, Hao; Yuan, Zheng-an

VERSION 1 – REVIEW

| REVIEWER | Antonio Sorlózano
University of Granada, Spain |
|---|---|
| REVIEW RETURNED | 08-Oct-2017 |

| GENERAL COMMENTS | I read with interest the paper by Gong and cols. It is a study conducted between May 2012 and May 2016 through stool samples collected by 22 hospitals in Shanghai. The manuscript contains interesting information about the epidemiology, clinical characteristics, etiology and seasonality of bacteria and viruses involved in infectious diarrhea in adults. The methodology is well described and could be useful for other Centers.
Minor changes:
• It would be desirable to include information on the detection of parasites that are also responsible for diarrhea (eg Giardia lamblia, Entamoeba histolytica…)
• Information should also be included regarding the immunological status of the patients (immucocompetent vs. immunosuppressed patients) |
|---|---|

| REVIEWER | Hans-Jörg Epple
Department of Gastroenterology, Rheumatology and Infectious Diseases, Charité – Universitätsmedizin Berlin, Campus Benjamin Franklin, Berlin, Germany |
|---|---|
| REVIEW RETURNED | 10-Dec-2017 |

| GENERAL COMMENTS | Using data from a municipial, hospital-based surveillance system, Gong and co-workers report on the microbial etiology of acute, non-epidemic infectious diarrhea in Shanghai. The microbial data are correlated with demographics such as well as with clinical symptoms of the patients.

The strengths of the study are the large sample of diarrheal patients included, the use of an established sentinel hospital-based surveillance system, and the broad spectrum of enteropathogens detected by appropriate microbial testing. The paper is the first report on the etiology of sporadic acute infectious diarrhoea in adults living in Shanghai. Analysis of stool samples obtained randomly from the included patients showed a high positivity rate for both viral and |
|---|---|

	bacterial enteropathogens. Compared to prior studies performed in Europe or the USA, the positivity rate for enteropathogenic bacteria was quite high (14.8%), and there was a different microbial spectrum with V. parahaemolyticus as the most prevalent species and Campylobacter playing a minor role.

Major points

• The results add little to the clinical management of patients presenting with acute diarrhoea.
• According to the manuscript, only patients with "mild diarrhea" were included into the study. It is well known that more severe courses of gastroenteritis are associated with a bacterial etiology. Exclusion of cases with more than mild symptoms can therefore bias the microbial results . The question arises for what reasons more severely ill patients were not included into the study and which definition of "non-mild" diarrhea was used to exclude subjects from the analysis? The definition of surveillance subjects provided in the methods section would include mild as well as severe cases.
• The presentation of the demographic data is confusing. As shown in table 1, the positive samples within the different age strata are given as absolute numbers and percentages ranging between 22.3 and 26.4%. Obviously the percentages were obtained by dividing the number of positive samples within a certain age stratum by the number of positive samples (n=4219; 100%) obtained from all strata. For comparison of the age-related prevalence, use of the positivity rate within each age-stratum (that is the number of positive samples divided by the number of positive and negative samples within the same age stratum multiplied by 100) seems more appropriate.
• The same consideration applies to the suburb data. Furthermore, a definition of suburb (as opposes to central Shanghai) would be helpful.
• The positivity rates found in different patient strata obviously depends on eating habits (exposure to potentially contaminated food) and on the fact whether or not the patients seek health care in the hospital. Therefore, the term "vulnerability" should be avoided.

Minor points:

• The sentinel hospitals comprising the surveillance system were chosen using a "Probability Proportionate to Size" sampling method. The method for the selection of stool within the sentinel hospitals is only roughly described. How were the sampling intervals chosen in the respective hospitals?
• The item "suspicious food" should be defined in more detail.
• Discussion, page 22 paragraph 2. The apparent lower prevalence of viral as compared to bacterial enteropathogens found in diarrheal patients in Africa might reflect the methods applied in the respective studies, many f which did not use PCR screening for the different stool viruses. In addition, the reviews cited (4 and 5) focus on diarrhea in children no adults.

In summary, the results presented are interesting from an epidemiological point of view. However, as the analysis of the date add little to the clinical management of patients presenting with acute diarrhoea, publication in an epidemiological journal might be more appropriate.

REVIEWER	William A. Petri, Jr.

	University of Virginia
REVIEW RETURNED	21-Dec-2017

GENERAL COMMENTS	Important addition to the literature for its description of the causes of diarrhea in adult outpatients in Shanghai. The only limitation is the use for the most part of classical methods (culture) for the identification of bacterial cases of diarrhea. Molecular testing is more sensitive.

REVIEWER	John Harris University of Liverpool, United Kingdom
REVIEW RETURNED	14-Mar-2018

GENERAL COMMENTS	Overall a reasonable paper, fairly clearly written but could do with some editing for style and language. The paper is a survey of hospital outpatients attending a number of hospitals for diarrhoeal disease. The participants enrolled into the study were sampled using a method that I am not familiar with but appears to be based on the size of the number of patients attending. I have a few comments for the authors to consider to improve the clarity of the paper in some places. Abstract, this could do with some editing particularly for language and style. The use of acronyms (e.g. PPS, DEC) before being written in full would aid clarity. Strengths and limitations, the authors refer to asymptomatic cases. I assume these are outpatients who have attended and reported symptoms of diarrhoea. In which case they are not asymptomatic but symptomatic cases but for whom no pathogen was detected. If they are asymptomatic then it begs the question what they were cases of. This term is used further in the text and I think it needs clarifying. Methods Laboratory tests A question on the detection of adenovirus, were these specifically group F (or adenovirus 40/41) because not all adenoviruses cause diarrhoea most cause respiratory disease and can still be excreted in stools. This could lead to the overestimation of the effect adenovirus has on diarrhoeal disease. Statistical analysis. It isn't clear to me how the models were derived when using the regression models. For example they use stepwise methods but it isn't clear how they decided to exclude variables from the model, although they state a two tailed pvalue was used was this level used to exclude the variables? Also they say nothing about how the decided on which was the more appropriate model when variables were excluded. I would also like more information on what variables they used as confounders or what variables the authors considered were confounders. Results Demographics and epidemiological characteristics The authors report significant differences firstly in the age groups between those where a pathogen was detected versus those where

they were not (I am assuming pathogen negative group are diarrhoea patients where no pathogen is detected). They haven't specified which test they used (Mantel-Haenszel?) to derive this statistic and I also wonder if this significance is due in part because of the number (and hence the power) in the sample. Also it isn't clear from this (or Table 1) which is leading to the significance. I feel that this would be better illustrated in a graph showing how (for example the age groups) differ between the two groups, furthermore, it would also be good to see how the age groups differ from the population of Shanghai and this would give a reader a better understanding of the representativeness of the study.
Table 2 adjusted odds ratio….
I can't follow how the adjustment has happened. I think this is better demonstrated with by illustrating the full model, and the coefficients observed and the final model, I can't see what the models are adjusting for, in the methods it refers to confounders but I can't see from this what they have included as confounders (or not). The first line of this table says male v female but all the others have their own line. I would have thought that only male or female need be displayed and indicating which the reference group is.
3 Clinical symptoms
It isn't clear what test has derived the p value for the comparison of symptoms between the positive and negative diarrhoea patients.

Figure 1 can be removed, it is difficult to read, figure 2 largely illustrates the same thing, if it is decided to keep both they should be enlarged to enhance their readability.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1

Thank you very much for your recognition.

Minor changes

• **It would be desirable to include information on the detection of parasites that are also responsible for diarrhea (eg Giardia lamblia, Entamoeba histolytica…).**

Thank you for your advice. Including information on detection of parasites involved in diarrhea patients, would improve the integrity of the pathogen spectrum. At the same time, because of limited financial resources and human resources, our surveillance only did the detection of bacteria and viruses now. In the future, we will try to include the detection of parasites in our surveillance.

• **Information should also be included regarding the immunological status of the patients (immucocompetent vs. immunosuppressed patients).**

Thank you for your advice again. Including information on immunological status of the patients would reduce the effect of confounders. At the same time, we haven't included this information in our

standardized questionnaire in the online system. In the future, we will try to include the information on immunological status to improve our surveillance.

■···

Reviewer #2

Thank you very much for your recognition.

Major points:

• The results add little to the clinical management of patients presenting with acute diarrhoea.

Thank you for your opinion. This study aimed to identify the epidemiology, clinical characteristics, etiology and seasonality of sporadic infectious diarrhea in adults in Shanghai. We focus more on etiology and epidemiology of diarrhea, and less on clinical management in this study. And we will do more research on clinical management of diarrhea patients based on this surveillance system. Thanks for your advice.

• According to the manuscript, only patients with "mild diarrhea" were included into the study. It is well known that more severe courses of gastroenteritis are associated with a bacterial etiology. Exclusion of cases with more than mild symptoms can therefore bias the microbial results . The question arises for what reasons more severely ill patients were not included into the study and which definition of "non-mild" diarrhea was used to exclude subjects from the analysis? The definition of surveillance subjects provided in the methods section would include mild as well as severe cases.

Thank you for proposing this question. I apologize for not make my point clearer. We did not exclude cases with more than mild symptoms. Surveillance subjects were defined as patients who visited the enteric disease clinics of sentinel hospitals, with 3 or more loose or liquid stools per day, or more frequently than normal for the individual. We meant that patients who visited the enteric disease clinics were not severe gastroenteritis patients who need to be hospitalized. So we mentioned the "mild" patients in "Strengths and limitations" and "Discussion". To avoid ambiguity, we revised this expression in Page 26. Now this sentence is "*Only diarrhea patients who visited the enteric disease clinics were included in surveillance, severe diarrhea patients or asymptomatic patients were possibly not studies in our research.*".

• The presentation of the demographic data is confusing. As shown in table 1, the positive samples within the different age strata are given as absolute numbers and percentages ranging between 22.3 and 26.4%. Obviously the percentages were obtained by dividing the number of positive samples within a certain age stratum by the number of positive samples (n=4219; 100%) obtained from all strata. For comparison of the age-related prevalence, use of the positivity rate within each age-stratum (that is the number of positive samples divided by the number of positive and negative samples within the same age stratum multiplied by 100) seems more appropriate.

Thank you for your constructive suggestion. We revised table 1 and related sentence accordingly in Page 13 and Page 12. And now the sentence is "*A significantly difference in positive rate within different age groups could be found among…*". Now the table, especially the comparison of the age-related prevalence, is more appropriate.

• The same consideration applies to the suburb data. Furthermore, a definition of suburb (as opposes to central Shanghai) would be helpful.

Thank you for your advice. We revised table 1 and related sentence accordingly in Page 13 and Page 16. And now the sentence is "*The positive rates in suburb and rural groups were significantly different (p<0.0001, Table 1). Comparing different enteric pathogen infections, the positive rates of patients in suburb and rural groups were significantly different (p<0.0001)*". Now the table is more appropriate.

We add an explanation of suburb or rural in "Methods" part in Page 10. The explanation is "*Patients who visited hospitals in suburb area were grouped in "suburb". Patients who visited hospitals in rural area were grouped in "rural"*". Really appreciate your suggestion.

• The positivity rates found in different patient strata obviously depends on eating habits (exposure to potentially contaminated food) and on the fact whether or not the patients seek health care in the hospital. Therefore, the term "vulnerability" should be avoided.

Thanks. The terms "vulnerable" has been deleted in the manuscript.

Minor points:
• The sentinel hospitals comprising the surveillance system were chosen using a "Probability Proportionate to Size" sampling method. The method for the selection of stool within the sentinel hospitals is only roughly described. How were the sampling intervals chosen in the respective hospitals?

Thank you for proposing this question. The question how sampling intervals were decided was complex. We comprehensively consider the hospitals' location (district distribution), classification (hospital level distribution) and annual number of diarrhea patients (workload and operability) to allocated different sample intervals to different hospital, ranging from 3:1 to 20:1. We revised accordingly in Page 7-8.

• The item "suspicious food" should be defined in more detail.

Thanks for your advice. The item "suspicious food" meant the suspicious food that may cause diarrhea, such as food which was contaminated by diarrhea pathogen. We added this sentence in "Methods" part in Page 10.

• Discussion, page 22 paragraph 2. The apparent lower prevalence of viral as compared to bacterial enteropathogens found in diarrheal patients in Africa might reflect the methods applied in the respective studies, many f which did not use PCR screening for the different stool viruses. In addition, the reviews cited (4 and 5) focus on diarrhea in children no adults.

Thank you for your correction. This sentence is deleted in the revised manuscript.

In summary, the results presented are interesting from an epidemiological point of view. However, as the analysis of the date add little to the clinical management of patients presenting with acute diarrhoea, publication in an epidemiological journal might be more appropriate.

Thank you very much for your recognition and earnest review. This study aimed to identify the epidemiology, clinical characteristics, etiology and seasonality of sporadic infectious diarrhea in adults in Shanghai. We hope than we can provide some points for prevention and control of infectious diarrhea in the perspective of preventive medicine, which is one branch of medicine. BMJ Open is dedicated to publishing medical research from all disciplines and therapeutic areas. So we submitted to BMJ open and sincerely hope our manuscript can meet your approval and is suitable for publication.

■ ■

Reviewer #3

• Important addition to the literature for its description of the causes of diarrhea in adult outpatients in Shanghai. The only limitation is the use for the most part of classical methods

(culture) for the identification of bacterial cases of diarrhea. Molecular testing is more sensitive.

Thank you for your recognition and encouragement. Molecular testing, no doubt, is more sensitive and better. However, because of limited financial resources, the identification of bacteria is more common and accessible in our surveillance work. We sincerely hope we can use molecular testing in the surveillance in the future.

■ ■

Reviewer #4

Thank you very much for your recognition.

• Abstract, this could do with some editing particularly for language and style. The use of acronyms (e.g. PPS, DEC) before being written in full would aid clarity.

Thank you for your correction and suggestion. We add full name of the acronyms in Page 3 accordingly.

• Strengths and limitations, the authors refer to asymptomatic cases. I assume these are outpatients who have attended and reported symptoms of diarrhoea. In which case they are not asymptomatic but symptomatic cases but for whom no pathogen was detected. If they are asymptomatic then it begs the question what they were cases of. This term is used further in the text and I think it needs clarifying.

Thank you for your comments. The word "asymptomatic" was used twice in the manuscript within the same sentence, in "Strengths and limitations" and "Discussion" parts. "*Only diarrhea patients who visited the enteric disease clinics were included in surveillance, severe diarrhea patients or asymptomatic patients were possibly not studies in our research.*" Surveillance subjects were patients who visited the enteric disease clinics of sentinel hospitals with diarrheal symptom. So we meant that asymptomatic patients were possibly not included in this study, which was one of our limitations. There may be misunderstood. Or I have not made this point clear. Hope to get your further guidance.

Methods

Laboratory tests

• A question on the detection of adenovirus, were these specifically group F (or adenovirus 40/41) because not all adenoviruses cause diarrhoea most cause respiratory disease and can

still be excreted in stools. This could lead to the overestimation of the effect adenovirus has on diarrhoeal disease.

Thank you for proposing this question. Yes, the detection of adenovirus in our surveillance was specially adenovirus 40/41 which cause diarrhea. We have replaced "adenovirus" with "enteric adenovirus" throughout the manuscript, which improves the clarity of our paper.

• Statistical analysis. It isn't clear to me how the models were derived when using the regression models. For example they use stepwise methods but it isn't clear how they decided to exclude variables from the model, although they state a two tailed pvalue was used was this level used to exclude the variables? Also they say nothing about how the decided on which was the more appropriate model when variables were excluded.

I would also like more information on what variables they used as confounders or what variables the authors considered were confounders.

Thank you for your comments. We use the logistic procedure in SAS 9.3 with the option "selection=stepwise" (default parameters) to perform the stepwise regression. We put age group, gender, suburb, season, consumption of suspicious food, medical history of enteric disease, history of suspicious water, history of contact with similar diarrhea patients, history of dining together, whether to leave the city go out of city, and whether to keep a pet into the initial model. The final model included the following variables: age group, gender, suburb, season, consumption of suspicious food, medical history of enteric disease, and whether to keep a pet. We add this sentence about the variables include in model in "Methods" part in Page 10.

Results

Demographics and epidemiological characteristics

The authors report significant differences firstly in the age groups between those where a pathogen was detected versus those where they were not (I am assuming pathogen negative group are diarrhoea patients where no pathogen is detected). They haven't specified which test they used (Mantel-Haenszel?) to derive this statistic and I also wonder if this significance is due in part because of the number (and hence the power) in the sample. Also it isn't clear from this (or Table 1) which is leading to the significance. I feel that this would be better illustrated in a graph showing how (for example the age groups) differ between the two groups, furthermore, it would also be good to see how the age groups differ from the population of Shanghai and this would give a reader a better understanding of the representativeness of the study.

Thank you for your suggestions and careful review. Yes. "Cochren-Mantel-Haenszel test was used for comparison of categorical variables.". To improves the clarity of our study, we add this sentence below table 1, table 3 and table 4 in Page 14, Page 19 and Page 22 respectively.

As to the age groups, we adjusted table 1. For comparison of the age-related prevalence, now we use the positivity rate within each age-stratum.

Table 2 adjusted odds ratio….
I can't follow how the adjustment has happened. I think this is better demonstrated with by illustrating the full model, and the coefficients observed and the final model, I can't see what the models are adjusting for, in the methods it refers to confounders but I can't see from this what they have included as confounders (or not). The first line of this table says male v female but all the others have their own line. I would have thought that only male or female need be displayed and indicating which the reference group is.

Thank you for your comments. According to the stepwise logistic procedure in SAS, the final model included the following variables: age group, gender, suburb, season, consumption of suspicious food, medical history of enteric disease, and whether to keep a pet. The adjusted odds ratio of a variable was the odds ratio within the final model adjusting for the other variables in the model.

3 Clinical symptoms
It isn't clear what test has derived the p value for the comparison of symptoms between the positive and negative diarrhoea patients.
Thank you for your comments. Yes. "Cochren-Mantel-Haenszel test was used for comparison of categorical variables.". To improve the clarity of our study, we add this sentence below table 3 in Page 19.

Figure 1 can be removed, it is difficult to read, figure 2 largely illustrates the same thing, if it is decided to keep both they should be enlarged to enhance their readability.

Thank you for your suggestions. We re-think your suggestions earnestly and sincerely hope that Figure 1 could be kept. The broad spectrum of enteropathogens is one of the strengths of our study. Figure 1 showed the broad pathogen spectrum well. It can also show the time trend of different enterpathogens and reflect the shift of the dominant role of bacterial infection or viral infection visually. Really hope figure 1 could be kept. Hope to get your further guidance.

VERSION 2 – REVIEW

REVIEWER	John Harris
	University of Liverpool, UK
REVIEW RETURNED	30-Apr-2018

GENERAL COMMENTS	The manuscript is an improvement on the previous version and the authors have responded appropriately to my previous comments.
	I have some additional comments that are easily rectified.
	The authors refer to suspicious food, can they expand on this, is this something that the person self reports or is it food taken from a list in the questionnaire.
	The authors could provide the questionnaire as an annex/appendix
	Figure 1 is still difficult to read.
	The manuscript requires some work for grammar prior to publication.

VERSION 2 – AUTHOR RESPONSE

Reviewer #4
Thank you very much for your recognition and constructive suggestions.

• The authors refer to suspicious food, can they expand on this, is this something that the person self reports or is it food taken from a list in the questionnaire.
Thank you for your correction and suggestion. We expand on the "suspicious food" accordingly.

• The authors could provide the questionnaire as an annex/appendix.
Thank you for your suggestion. The standardized questionnaire we used was designed according many public health workers' epidemiological practice. Considering the intellectual property and related issues, we have no rights to provide the questionnaire in this manuscript. We are deeply sorry about this. The questionnaire includes demographic, epidemiological and medical information. If you have any other questions about the questionnaire in our study, we could add words to explain accordingly.

• Figure 1 is still difficult to read.
Thank you for your comment sincerely. The Figure 1 in peer review PDF is so small that it is difficult to read. We don't know if you could see the figure 1 file. If you could, I think it will interest you. In regard to Figure 1, we hope it can not only demonstrate the time trend and the shift of the dominant role trend of bacterial diarrhea and viral diarrhea periodically, but also roughly reflects the broad pathogen spectrum, the main bacterial pathogens and viral pathogens. We believe the fourth point of results (explanation of the seasonality of pathogen spectrum), Figure 1 and Figure 2 together can adequately and intuitively demonstrate our pathogen spectrum. Hope to discuss with you and get your further guidance.

• The manuscript requires some work for grammar prior to publication.
Thank you for your suggestion. We check writing throughout the manuscript and try to improve the grammar as much as possible. Hope now the manuscript meets the standard.