Author's response to reviews

Title: Combined Mathematical Model in predicting incidence of hepatitis E using 2000-2012 morbidity data from Shanghai, China

Authors:

Hong Ren (hren@scdc.sh.cn)
Jian Li (jli_2@scdc.sh.cn)
Zheng-An Yuan (zayuan@scdc.sh.cn)
Jia-Yu Hu (jyhu@scdc.sh.cn)
Yan Yu (yuyan08@yahoo.com)
Yi-Han Lu (luyihan@fudan.edu.cn)

Version: 2 Date: 2 August 2013

Author's response to reviews: see over
Dear Philippa Harris:

We appreciate the advice from Miss Sheryl Ramos and the valuable comments from the reviewers. We have addressed their comments as below in italics and marked any changes highlighted with the color of red.

Editorial Requirement:

(* ) Copyediting:
After reading through your manuscript, we feel that the quality of written English needs to be improved before the manuscript can be considered further. We advise you to seek the assistance of a fluent English speaking colleague, or to have a professional editing service correct your language. Please ensure that particular attention is paid to the abstract.

**Responses:** We have revised the manuscript with the assistance of a native English speaker with scientific expertise.

Additionally, please clarify who gave permission to access the database and if any ethical approval was required. Please include this information in the methods section.

**Responses:** We have clarified the permission of database access and declared no ethical issues in the Methods section.

Please also ensure that your revised manuscript conforms to the journal style (http://www.biomedcentral.com/info/ifora/medicine_journals). It is important that your files are correctly formatted.

**Responses:** We have complied with the journal style including both format and spellings.

Reviewer: Tassanee Silawan

Overall

The paper should be properly checked for English corrections before being published

**Responses:** We have revised the manuscript with the assistance of a native English speaker with scientific expertise.

Abstract

Background:

‘Therefore, some improved tools based on mathematical mode were developed to forecast
hepatitis E epidemics.”

Do the authors want to forecast “epidemics” or “incidence”? Methods for forecast epidemics are different from the methods for forecast incidence.

Reponses: We have revised “mode” and “epidemics” to “model” and “incidence”, respectively.

“Due to both linear and nonlinear patterns of time series morbidity data of infectious disease such as hepatitis E, applicable statistical model cannot be easily determined.”

This should directly mention only for hepatitis E, focus on the reasons for using combined mathematical model.

Reponses: We have revised the sentence accordingly.

Methods:
The following details should be added:
- Source (surveillance system?) and characteristics (laboratory confirmed?) of data
- Data points (units) for analysis (Annual/ Monthly/ Cases/ Rates?), - Explain more about (and how to develop) the combined model
- Time horizon for training (fitted) set and testing (predicted) set for model diagnostic check
- Time horizon for fitted data and forecasted data

Reponses: We have added the detailed information in the Abstract and Methods section.

Results:
- The equation of ARIMA would be skipped, few people can understand and we rather want to know the equation of the combined model…. of single ARIMA model and combined model….
- Specify the combined model

Reponses: We have deleted the equation of ARIMA. Considering neuron model and network architectures of BPNN were based on the black-box theory, we cannot specify an equation of ARIMA-BPNN combined model, so we have illustrated Figure 1 to show how to combine ARIMA and BPNN.

The predicted incidence of hepatitis E in 2013 would keep fluctuating within a narrow range from 0.095 to 0.372 per 100 000 persons with a seasonal variation.

The word “predicted” …….revise to “forecasted”

- Clearly explain about the seasonal variation

Reponses: We have revised “predicted” to “forecasted”. Also, we have added the details describing the seasonal variation in the Abstract and Results section.

Conclusions:
- Conclusions should not beyond the findings.

Reponses: We have revised them accordingly.

Background
The last paragraph:
Add evidences or findings from previous researches which indicated an inaccuracy (not so fit) of those models in the last paragraph.

Responses: We have added the findings of two previous researches in the last paragraph of Background to support our opinions.

Materials and Methods
1st paragraph:
-Explain how to calculate the monthly rates. Was the same population a denominator for all 12 months in the same year?

Responses: In our study, monthly morbidity rates of hepatitis E was calculated with the total number of new cases in one month as numerator, and annual average population in Shanghai as denominator.

Results
Should explain more details of the results. The focus should be on comparison between ARIMA and ARIMA-BPNN combined model.

Responses: We have added more details of ARIMA model and ARIMA-BPNN combined model in the Results section.

Data description
-It should be “General patterns of hepatitis E”
-Graphic presentation (line graph of monthly rates) should be presented instead of table. The characteristics of trend and seasonal variations should be clearly explained.

Responses: We have revised the sub-heading to “General patterns of hepatitis E”. Additionally, we have added Figure 2 to illustrate the trend and seasonal variations of hepatitis E morbidity in Shanghai; however, we prefer to save Table 1 due to the reason that it contained more information.

“The male morbidity was significantly higher than that of female (t = 8.951, P < 0.001)”
-What were the reasons for comparison of morbidity between male and female? Were these differences taken into account in forecast?

Responses: In our study, the morbidity between male and female was compared only to explain the epidemiological characteristics of hepatitis E in the Chinese population, which was not further taken into account in forecast.

“Meanwhile, X-12-ARIMA seasonal adjustment program showed that the monthly morbidity of hepatitis E had seasonal and trend variations (F = 40.02, P < 0.001)”
-No evidence support.

Responses: The X-12-ARIMA seasonal adjustment package developed by the United States Bureau of the Census is comprehensive and available for seasonal adjustment to each individual series (i.e. trend-cycle, seasonal component, and irregular component). In our study, we used the X-12-ARIMA seasonal adjustment program of Eviews to explain whether there were seasonal or trend variations in the original monthly morbidity data of hepatitis
E. And the results of F test have confirmed the stable seasonality. Related information was presented in the Reference 21.

The best-fitting ARIMA model
- Is logarithmic transformation needed?

Reponses: Based on the repeated Augmented Dickey-Fuller tests, we thought that the logarithmic transformation was necessary for stabilizing the monthly morbidity series of hepatitis E.

- The following results should be added and clearly explained;
  • error of the fit model
  • predicted values (from model identifying), and how the predicted values close to the actual values.
  • forecasted values in 2013 (line graph by month should be presented).

Reponses: We have added the parameters explaining the error of the fit model and the description of how the predicted values from best-fitting ARIMA model was identified close to the actual ones. Additionally, we have added Figure 2 to present the actual values, predicted values in 2012, and forecasted values in 2013 from single ARIMA and ARIMA-BPNN combined model, respectively.

ARIMA-BPNN combined model
- should explain on predicted values (from model identifying), and how the predicted values close to the actual values.
- Line graph by month should be presented (can be compared with ARIMA).

Note: If possible, some of the actual data in year 2013 should be compared with the forecasted values, both ARIMA and ARIMA-BPNN combined model.

Reponses: Similar to last response, we have added the description of how the predicted values from ARIMA-BPNN model was identified close to the actual ones. Additionally, we have added Figure 2 to present the comparison of predicted and forecasted values between single ARIMA and ARIMA-BPNN combined model. However, the actual data in 2013 has not been available so far.

Discussion
“Compared with other infectious diseases, the epidemic of hepatitis E is influenced by individual, environmental and social factors”

- Other infectious diseases are also influenced by those factors.

Reponses: We have revised the comments in the sentence to indicate that compared to blood-borne infectious disease (e.g. hepatitis B, hepatitis C, AIDS), hepatitis E is more affected by environmental and natural factors which lead to more observable seasonal variation of morbidity.

“These multiple factors will lead us to the difficulties attempting to employ variables regression models to forecast the epidemics. Time series analysis method takes advantage of associations in the seasonally and sequentially lagged relationships. Without independent variables, it can also automatically determine the best-fitting model.”
These are not reasonable.

**Responses:** We have revised the comments accordingly to suggest that time series analysis has the advantage of forecasting the morbidity without focusing on specific risk factors.

The discussion should cover the following key findings of the current research by comparing with the previous research and giving the reasons for those phenomenon or findings.
- Pattern; trend and seasonal
- Model, and Forecast.
- The effectiveness of the model; strength, weakness, points need to be concern
- Implications

**Responses:** We have added these key comments in the Discussion section.

**Conclusions**
- Conclusions should not beyond the findings.

**Response:** We have revised them accordingly.

**Table 1**
- Should present monthly cases or rates (per 100,000 pop.) of…from..to… using line graph (to show the general patterns, trends and seasonality)

**Responses:** We have revised Table 1 as well as added Figure 2 to illustrate the general patterns, trend and seasonality of morbidity of hepatitis E from 2000 to 2012 in Shanghai.

**Table 2**
- Can be omitted. The significant estimated parameters should be > 2.0.

**Responses:** We prefer to save Table 2 because it contained the parameters of the equation of best-fitting ARIMA model.

**Table 3**
- Forecasted rates from Jan to Dec 2013 should be separated from table 3 and better presented using line graph to show the magnitude and seasonality
- Table 3 Predicted and error rates of …..from….to…..
- Date …..revise to Month, -Morbidity (per 100,000) …..revise to Morbidity (per 100,000 pop.)
- Validation…..revise to Predicted rate
- Des 2012…..revise to Dec 2012, -The last line should present the mean error rate (MER).

The following line graphs should be presented.
- Line graph of monthly cases or rates from…..to…….(to show general patterns, trend, and seasonal of the disease).
- Line graph of monthly cases or rates from…..to……( to show forecasted values).

**Responses:** We have revised Table 3 and added Figure 2 to present the monthly morbidity rates, predicted and forecasted values.

**Discretionary Revisions**
To make the article more attractive, the words “using 2000-2012 morbidity data from”
should be removed from the title. The title may be
• Combined Mathematical Model in predicting incidence of hepatitis E in Shanghai, China. Or
• Combined Mathematical Model to forecast hepatitis E incidence in Shanghai, China. Or
• Forecast for hepatitis E incidence in Shanghai of China using Combined Mathematical Model.

Responses: We have revised the title to “The development of a combined mathematical model to forecast the incidence of hepatitis E in Shanghai, China”.

Reviewer: Naiqing Zhao
1. Authors should give the final forecasting expression

Responses: Considering neuron model and network architectures of BPNN were based on the black-box theory, we cannot specify an equation of ARIMA-BPNN combined model, so we have illustrated Figure 1 to show how to combine ARIMA and BPNN.

2. Prediction and forecasting is different. Forecasting is often used in time series analysis.

Responses: We have revised “prediction” to “forecasting” in the manuscript, except for the “prediction” being used in model fitting values in 2012.

3. The concepts of morbidity and incidence may not be the same. "morbidity" may be applicable to the whole manuscript.

Responses: We have revised “incidence” to “morbidity” in the manuscript, except for the “incidence” being used in the forecasting of hepatitis E in 2013.

We hope that we have addressed all comments to your satisfaction.

Sincerely yours,

Yan Yu, MD
Associate Professor of Public Health
Shanghai Municipal Center for Disease Control and Prevention

Yi-Han Lu, MD, PhD
Lecturer of Epidemiology
Fudan University School of Public Health