Author’s response to reviews

Title: Forecasting incidence of hand foot and mouth disease using BP neural networks in Jiangsu province, China

Authors:

Wendong Liu (jscdclwd@sina.cn)
Changjun Bao (Bao_2000@163.com)
Yuping Zhou (zyp@jscdc.cn)
Hong Ji (719494589@qq.com)
Ying Wu (52236007@qq.com)
Yingying Shi (670129037@qq.com)
Wenqi Shen (m15250954634@163.com)
Jing Bao (839259894@qq.com)
Juan Li (591564075@qq.com)
Jianli Hu (59139316@qq.com)
Xiang Huo (63093305@qq.com)

Version: 1 Date: 29 May 2019

Author’s response to reviews:

Dear editor,

Thank you very much for your letter of 26 Apr 2019, in which you encouraged us to revise our manuscript “Forecasting incidence of hand, foot and mouth disease using BP neural networks in Jiangsu province, China” (Manuscript ID: INFD-D-19-00160).

The reviewers’ insightful comments enabled us to greatly improve the quality of our manuscript. According to the reviewers’ valuable comments, we have some significant revisions on the manuscript. Enclosed please find the copy of our revised manuscript with all changes track marked.
The following is the point-to-point response to the reviewers’ comments. We hope that the revisions and our accompanying responses will be sufficient to make our manuscript suitable for publication in BMC Infectious Diseases.

Yours sincerely,

Wendong Liu
Jiangsu Provincial Center for Disease Control and Prevention
172, Jiangsu Rd.
Nanjing 210009
China
Tel:+862583759472
Fax:+862583759549
Email:jscdclwd@sina.cn

Response to Reviewer 1.

Authors have developed a back propagation neural networks (BP model) to predict the future trend of HFMD in Jiangsu province, China. However, the paper doesn't present any significant outcome.

Response: Dear reviewer, many thanks for this comment. Hand, foot and mouth disease (HFMD) is a rising public health problem worldwide, and effective prevention and control of HFMD has become a major challenge in the field of public health. In this study, we constructed a multivariate BP model which comprehensively combined the autocorrelation of the independent, the climatic variables and their hysteresis effects. It performed satisfactorily in predicting the HFMD epidemic. This model has a good prospect of practical application. Its use in the detection and prediction of HFMD may provide an opportunity for re-allocating healthcare resources. To our knowledge, it is the first study to develop BP model with climate factors for predicting HFMD. Particularly, we proved that climate factors could significantly improve the prediction accuracy of the BP model. And BP model is more suitable for HFMD prediction than ARIMA model in the study area. We believe our findings not only has practical application value, but
also can be used as beneficial reference for other similar researches. It would appeal to a broad audience including those in public health and scientific communities.

Your comments are very valuable for us to improve the manuscript. In this version, we have made some major revisions. We sincerely hope it will gain your approval.

Specific comments

1. Abstract is not clearly expressing the idea. Sentences need to be reframed
Response: Thanks for your comment. We have rewritten the abstract.

2. Introduction
* Paper describes only EVA71 and CVA16 as an etiology of HFMD. What are the other enteroviruses which cause HFMD in China? (line 53,54 and 55)
Response: Thanks for your comment. We have added information about other enterovirus as: recently, enteroviruses other than EV-71 and CV-A16 has been increasing in both mild and severe cases and Coxsackie virus A6 (CV-A6) has been emerging as another predominant serotype in some regions.

* Considerable attention shall be given worldwide prevalence and incidence of HFMD (line 58-59)
Response: Thanks for your suggestion. We have briefly described the worldwide prevalence of HFMD with some references: It is a rising public health problem and has attracted considerable attention worldwide. It is especially widespread in Asia-Pacific areas and presents a general increasing incidence in recent decades.

* How to effectively prevent and control HFMD has become a major challenge in the field of public health (line 60-61)-incomplete sentence
Response: Thanks for your comment. It has been revised as: Effective prevention and control of HFMD has become a major challenge in the field of public health.
Recently, some researchers are interested in forecasting the incidence of HFMD, using the linear time series models such as auto-regressive integrated moving average (ARIMA) (line 68-69)- vague idea, it shall be specific.

Response: Thanks for your comment. This sentence has been revised as: Recently, some researchers are interested in forecasting the incidence of HFMD, using the liner time series models. For example, an ARIMA(1,0,1)(0,1,0)12 model was constructed to forecast the HFMD incidence in Sichuan, China. In another study, multivariable ARIMA models using search engine query data and climate factors as exogenous variables were developed to predict the HFMD epidemic in Guangdong, China.

3. Materials and Methods

* ARIMA model- how can it be used for forecasting incidence of HFMD? Concept is not well understood (line 104)

Response: Thanks for your comment. In this section, we added: “Since the incidence time series of HFMD commonly shows significant cyclical and seasonal patterns, auto-regressive integrated moving average (ARIMA) model should be considered to erect the benchmark model. Once the univariate ARIMA model was selected, the multivariate ARIMA model including climate factors as external regressors was further developed. In this study, ARIMA model that incorporates climate factors was referred as ARIMAX.”

* Model evaluation

The data between 2009 and 2014 were used as training set to fit models, and data between 2015 and 2016 were used as testing set to evaluate the forecasting accuracy of different models (line 126,127, 128)- how did you select training set and testing set? Any criteria applied? If you do not take enough data for testing set, how can you take it forward for analysis?

Response: Thanks for your comment. In this study, 96 months' data (2009.1-2016.12) were used, the first 72 months' data (2009.1-2014.12) were used to fit the model, and the last 24 months data (2015.1-1206.12) were used to evaluate the prediction effect of the model. The data in testing set accounts for 25% of the total data, it is sufficient to evaluate the prediction accuracy of the models.
4. Statistics analysis

* R version 3.5.0 was employed to explore the temporal characters of HFMD and conduct the spearman analysis between HFMD incidence and meteorological factors. BP neural models were constructed in Python version 3.7.0 and ARIMA model was fitted in SPSS version 20.0 (line 137-140)- Why did you chose multiple software for different model construction?

Response: Thanks for your comment. We have redone the data analysis just using R software, and this section was rewritten as: “all statistical analyses were completed using R software version 3.5.0. Particularly, ARIMA models were performed with R package “forecast” version 8.5. Meanwhile, BP models were constructed with R package “nnet” version 7.3-12.“

Notably, due to the randomization of initial weights, the results of BP models produced by R software were slightly difference from that by Python, which made no difference to the conclusion of this study.

5. Results

* Residual ACF and PACF of model ARIMA. (Line 194)- not understood to the reader

Response: Thanks for your comment. It has been deleted due to changes to the relevant contents.

* ARI, MAI and seasonal ARI(table line 196)- expand the abbreviations

Response: Thanks for your comment. This table has been deleted due to changes to the relevant contents.

6. Discussion

* Line 218-222- idea is not clear, meaning is missing

Response: Thanks for your comment. We have revised these sentences as following: As HFMD is a common infectious disease throughout the world, modeling its epidemic has been concerned and actively studied in recent years. Some researchers have put forward different prediction methods for HFMD. For examples, Yu et al, developed a new hybrid model with ARIMA and nonlinear auto-regressive neural network. Zhong et at, employed XGBoost, one of machine learning methods, to forecast HFMD with multiple environmental factors.
* Line 267-273- Limitations of the study must be re written- generalized way of explanation not concentrating of this study.

Response: Thanks for your comment. We have revised this section.

* No significant outcome presented.

Response: Thanks for your comment. In this study, we constructed a multivariate BP model using climate factors which performed satisfactorily in predicting the HFMD epidemic. This model has a good prospect of practical application. Particularly, we proved that climate factors could significantly improve the prediction accuracy of the BP model. And the BP model is more suitable for HFMD prediction than ARIMA model in the study area.

Response to Reviewer 2.

Main issues:

Methods/experimental design: For a fair comparison between two models, the ARIMA should also consider various climate covariates in modeling. Thus, based on current manuscript, I am not convinced by the results "ARIMA model performed the worst in the three models. It could not be used to predict the real data, with a MAPE as high as near to 40%" in Line 41-43. Also, be careful to use the second conclusion that is a very strong judgment.

Response: Thanks for your comment.

(1) Based on the selected univariate model ARIMA(1,0,1)(1,1,0)12, we fitted 15 multivariate ARIMAX models with different climate factors. The model with one order lagged temperature as external predictor was the optimal ARIMAX model, with a minimum AIC=1132.37 and a minimum BIC=1142.76. However, this model did not performed much better than the univariate ARIMA model, its MAPE was similar to that of the ARIMA model. Notably, both the ARIMAX model and the ARIMA model performed much worse than the BP models. See table 3 and table 4.

(2) On the second question, we have revised it as “It could not be used to predict the HFMD incidence in the study area, with a MAPE as high as near to 40%.”

Moreover, for Figure 3 and table 2, I highly doubt that this parameter specification (101,110) of ARIMA model from reference [12] is also reprehensive or best suitable for the HFMD case of this work, as they used totally different data within different study area.
Response:

Response: Thanks for your comment. We have fitted eight alternative ARIMA models which were all statistically significant. The model ARIMA(1,0,1)(1,1,0)12 had the minimum AIC and BIC, so it was selected as the best-fitting ARIMA model to the time series of HFMD in Jiangsu province, China. See table 2.

Background section (Line 67-77) lacks necessary literature review, for instance: (1) as I know, many HFMD studies have considered different climate covariates in time series analysis, even in ARIMA model. What kinds of climate factors did they utilize? (2) Is this work the first study to apply BP model for HFMD time series analysis? If not, how did they specify BP model, with or without considering climate factors?

Response: This is the first study to apply BP model for HFMD prediction. We have added this idea in this section as: “To date, however, there has been no literature report on using BP model to predict the epidemic of HFMD.” In the discussion section, the prediction methods of hand foot and mouth disease and the ARIMA model with climate factors have been discussed. To avoid repetition, no additional literature review on HFMD prediction methods was added in the background section.

Section "BP neural networks" (line 93-103): due to that BP is the final selected model, only one short paragraph herein to introduce BP is totally not enough, especially when compared with next section "ARIMA model" in which authors introduced mathematical formulas. In addition, referential index (line 93 and 104) in sub-level titles seems not correct for publication.

Response: Thanks for your comment. This section has been rewritten. We briefly describe the BP model training process and present the main calculation steps. Besides, we re-insert the reference.

Results: the authors need to give the final model outputs of two climate factors (temperature and rainfall) and their lag effects, which is a vital finding for HFMD control and prevention in this study.

Response: Thanks for your comment. The BP model can establish the nonlinear mapping relationship between meteorological factors and the incidence of hand foot and mouth disease through self-organizing learning ability, so that it can achieve the purpose of accurate prediction. But due to the methodological limitation, it cannot evaluate the correlations between the climate factors and the HFMD incidence.
The conclusion (line 277-282) section need to be much improved to give more useful information focusing on the HFMD finding in this paper.

Response: Thanks for your comment. We have rewritten this section as follows: In this study, four models were constructed to forecast the incidence of HFMD in Jiangsu province, China. The BP models performed much better than the ARIMA models. The introduction of mean temperature, rainfall and their one order lagged terms significantly improved the prediction accuracy of the BP model. On the contrary, neither the univariate ARIMA model nor the multivariate ARIMAX model achieved satisfactory prediction accuracy. The climate factors did not optimize the performance of the ARIMA model. In general, the multivariate BP model comprehensively combined the autocorrelation of the independent, the climatic variables and their hysteresis effects. It is an ideal method to predict the HFMD epidemic, which has a good prospect of practical application.

Minor revisions

Be consistent of terminology: hand foot and mouth disease (line 49, 285) or hand, foot and mouth disease (line 25, line 53) or hand, foot, and mouth disease.

Response: Thanks for your comment. The unabbreviated disease name appeared everywhere except the reference section in the manuscript has been consistently rewritten as “hand, foot and mouth disease”.

Line 163: the sub-level title here seems not appreciate, may be "statistical software"

Response: Thanks for your comment. We have revised this sub-title as “Statistical software”.

Line 165-167: The authors should clearly introduce how they selected these eight factors (x1-x8) before or within this section, such as why using 1-4 months (x1-4), and what types of lag effects (x7 and 8).

Response: Thanks for your comment.

(1) The reason for selecting x1-x4 was due to the autocorrelation analysis, which was described as: According to the results of autocorrelation analysis, the time series of the dependent variable presented significant autocorrelation at lag 1-4. Given this, the four lagged terms were considered as predictors in the BP model.
(2) The reason for selecting x5-x6 was clarified with the univariate spearman correlation analysis.

(3) The reason for selecting x7-x8 was added as: Further cross-correlation analysis indicated that both mean temperature and rainfall significantly related with the incidence of HFMD at lag 1, with correlation coefficients of 0.235 (p=0.0216) and 0.251 (p=0.0146) respectively.

Line 36: "both in" to "in both"
Response: Thanks for your suggestion. We have revised it.

Line 47: "important" to "importance"
Response: Thanks for your suggestion. We have revised it.

Line 245: "So" to "so"
Response: Thanks for your suggestion. We have revised it.