

The analysis of GM (1, 1) grey model to predict the incidence trend of typhoid and paratyphoid fevers in Wuhan City, China

Xiaobing Yang, PhD^a, Jiaojiao Zou, MD^a, Deguang Kong, MD^a, Gaofeng Jiang, PhD^{b,*}

Abstract

Typhoid and paratyphoid fevers (TPF), systemic emerging infectious diseases, is a serious health problem for society. If the incidence trend of TPF can be predicted, prevention and control measures can be taken in advance to reduce the harm to the people's health.

Grey Model First Order One Variable [GM (1, 1)] was applied to predict the incidence trend of TPF with the incidence data of TPF in Wuhan City of China from 2004 to 2015. The original data were acquired from the national surveillance system.

The GM (1, 1) model was established as $y(t+1) = 0.88 e^{-0.21t} + 0.15$. The goodness-of-fit test indicated that the precision (degree 2) was qualified ($C=0.40$, $P=.91$). We further compared actual values with predicted values in 2016 and found that GM (1, 1) model we built has excellent performance in incidence trend prediction.

Our prediction shows that the TPF incidences in Wuhan City will be slowly decreasing in the next 3 years. It is, however, still necessary to strengthen the comprehensive prevention and control to reduce the incidence level of TPF.

Abbreviations: CISDCP = China Information System for Diseases Control and Prevention, GM (1, 1) = Grey Model First Order One Variable, TPF = typhoid and paratyphoid fevers.

Keywords: GM (1, 1) grey model, grey system theory, incidence, prediction, typhoid and paratyphoid fevers

1. Introduction

Typhoid and paratyphoid fevers (TPF), systemic emerging infectious diseases, are caused by *Salmonella enterica* serotype Typhi or serotype Paratyphi (A, B, or C), respectively.^[1] With improvements in municipal drinking water treatment, sanitation, hygiene, and food production and preparation, illness and death caused by TPF became rare in industrialized nations in Europe and North America, but it remains a serious public health problem in developing countries.^[2–4] The incidence of TPF in China has been gradually decreased and remained at a comparatively low level since 2004. However, TPF is still one of the important sporadic intestinal infectious diseases in Wuhan City.^[5–7]

Ingestion of contaminated water and food is the most common route of TPF transmission.^[8] In addition, many social factors, especially economic development level, health facilities, environmental factors, and living conditions can influence the incidence

of TPF.^[9–12] Environmental factors, such as climate, have also been investigated to assess their influence on water-/food-borne infections.^[13–16] As a result, effective preventive strategies are required to control TPF.

In 1982, Deng^[17] firstly established the grey system theory, which shows great capability for studying uncertainty problems with small sample, poor information, uncertain system, and lack of data. Grey system theory is developing from information theory, cybernetic theory and mathematical method to solve incomplete and uncertain problem, and it also accord with current system science and uncertain system theory. Grey system theory focuses on the “poor information” systems with “partial information known, partial information unknown.”^[18] The theory studies and forecasts the unknown area to master the whole system, through extracting valuable information from known information.^[19] During the last 3 decades, the grey system theory has been developed rapidly and caught the attention of many researchers. It has been widely and successfully applied in many fields such as social, scientific and technological, geological, and medical systems.^[20–23]

Although various types of grey models can be mentioned, most of the previous researchers have focused their attention on Grey Model First Order One Variable [GM (1, 1)] models in their predictions because of its computational efficiency. GM (1, 1) type of grey model is the most widely used in the literature, pronounced as “Grey Model First Order One Variable.”^[24] GM (1, 1) model is a time series forecasting model, which is able to make accurate predictions for forecasting of the monotonous type of processes. The differential equations of the GM (1, 1) model have time-varying coefficients. In other words, the model is renewed as the new data become available to the prediction model.

In this study, a GM (1, 1) model was proposed to make prediction of the TPF incidence trend based on the epidemiologic data of TPF in Wuhan City, and provide reference for the government in policy making.

Editor: Leyi Wang.

XY and JZ contributed equally to this work.

The authors of this work have nothing to disclose.

^a Wuhan Centers for Disease Prevention and Control, ^b Center for Translational Medicine, Tianyou Hospital, Wuhan University of Science and Technology, Wuhan, Hubei, China.

* Correspondence: Gaofeng Jiang, Center for Translational Medicine, Tianyou Hospital, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China (e-mail: jianggaofeng@wust.edu.cn).

Copyright © 2018 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Medicine (2018) 97:34(e11787)

Received: 23 October 2017 / Accepted: 14 July 2018

<http://dx.doi.org/10.1097/MD.0000000000011787>

2. Materials and methods

2.1. Data collection

The monthly case number of TPF and the average monthly incidence rates from 2004 to 2015 in Wuhan City were collected from the national surveillance system. TPF are notifiable diseases in China. All clinical and hospital doctors are required to report TPF cases to the local Center for Diseases Control and Prevention. In China Information System for Diseases Control and Prevention (CISDCP), both clinically and laboratory diagnosed TPF cases were collected without further distinction between typhoid and paratyphoid. Therefore, the clinically diagnosed and laboratory confirmed cases of typhoid fever and paratyphoid fever were all combined as TPF in CISDCP and in this study.

2.2. Ethics statement

The ethics committee of Wuhan Center for Disease Prevention and Control approved this study.

2.3. GM model principles

GMs predict the future values of a time series based only on a set of the most recent data depending on the window size of the predictor. It is assumed that all data used in grey models are positive, and the sampling frequency of the time series is fixed. In grey systems theory, GM (n, m) denotes a grey model, where n is the order of the difference equation and m is the number of variables.

In grey system theory, a grey prediction model is one of the most important parts, and the GM (1, 1) model is the core of grey prediction. The purpose of GM (1, 1) model is to work on system forecasting with poor, incomplete, or uncertain messages. The GM (1, 1) model has more advantages over those traditional prediction ways, because it does not need to know whether the prediction variables obey normal distribution, and also does not require too much statistic sample.

In order to smooth the randomness, the primitive data obtained from the system to form the GM (1, 1) is subjected to an operator, named Accumulating Generation Operator. The differential equation [ie, GM (1, 1)] is solved to obtain the n -step ahead predicted value of the system. Finally, using the predicted value, the Inverse Accumulating Generation Operator is applied to find the predicted values of original data.

2.4. Construction of grey prediction model GM (1, 1)

Let original series $x^{(0)}(i) = x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)$. By defining $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k=1,2,3, \dots, n$, we get a new series $x^{(1)}(k) : x^{(1)}(k) = x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)$. To some processes, $x^{(1)}(k)$ is the solution of the following white-formed ordinary differential equation.

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \tag{1}$$

where a and u are gray number, that is, pendent parameters, which are estimated by least square method. The equation (1) is called GM (1, 1). The solution of (1) is:

$$x^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{u}{a} \right] e^{-ak} + \frac{u}{a} \tag{2}$$

The equation (2) is called time response function. For $k \geq 2$, $x^{(0)}(k) = x^{(1)}(k+1) - x^{(1)}(k)$ is called predicting formula.

2.5. Accuracy testing of GM (1, 1)

Model with preferable fitting accuracy can be used to extrapolate predicted value. Otherwise, residual correction has to be carried out first. Usually, posterior error detection method is used to test the accuracy of GM (1, 1). The indexes of fitting testing includes posterior error ration (C) and small error probability (P).

Posterior error ration (C) is the ratio of residual standard deviation (S_e) and data standard deviation (S_x). Obviously, if the residual standard deviation is smaller, the prediction accuracy is more excellent. The specific formula is as followed:

$$C = \frac{S_e}{S_x} \tag{3}$$

In the formula (3),

$$S_e = \sqrt{S_e^2} \text{ and } S_e^2 = \frac{1}{N} \sum_{j=1}^N [e^{(0)}(j) - e]^2$$

$$S_x = \sqrt{S_x^2} \text{ and } S_x^2 = \frac{1}{N} \sum_{j=1}^N [x^{(0)}(j) - x]^2$$

Small error probability (P) is:

$$P = P\{|e^{(0)}(K) - \bar{e}| < 0.6744S_x\} \tag{4}$$

The prediction level is graded to 4 according to the 2 indexes above (Table 1).

2.6. Extrapolated prediction

If the result of fitting test is satisfactory, the model can be considered as credible. Then predicted value can be extrapolated by the followed formula:

$$\hat{x}(t) = \hat{y}(t-1), t = n+1, n+2, \dots, n+k \tag{5}$$

2.7. Statistical analysis

Excel 2013 was used to set up the database and establish predicting model for TPF incidence.

Table 1

Judgment standard of prediction accuracy for GM (1, 1).

Prediction accuracy grades	Small error probability (P)	Posterior error ration (C)
1 Excellent	$P \geq .95$	$C \leq 0.35$
2 Qualified	$.95 > P \geq .80$	$0.35 < C \leq 0.50$
3 Barely qualified	$.80 > P \geq .70$	$0.50 < C \leq 0.65$
4 Unqualified	$.70 > P$	$0.65 < C$

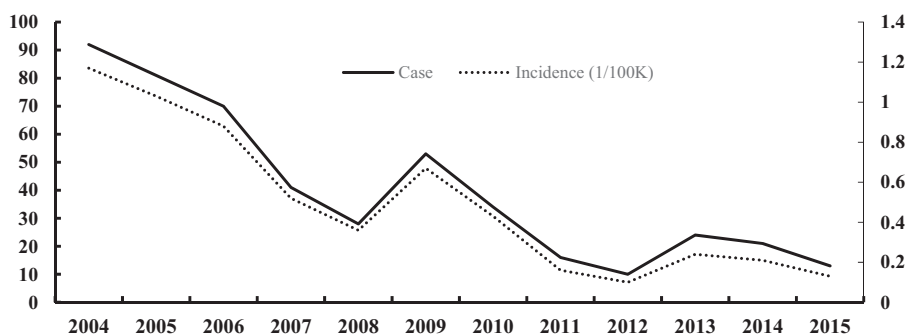


Figure 1. Typhoid and paratyphoid fevers incidences from 2004 to 2015 in Wuhan City of China.

3. Results

3.1. TPF incidences from 2004 to 2015 in Wuhan City

The incidences of TPF in Wuhan City from 2004 to 2015 are shown in Figure 1.

3.2. Establishment of GM (1, 1) model for TPF incidence prediction

Actual reported incidences of TPF from 2004 to 2015 in Wuhan were taken as time series to build GM (1, 1) predicting model. Its functional equation is as followed:

$$\hat{y}(t + 1) = 0.88e^{-0.21t} + 0.15 \tag{6}$$

3.3. Accuracy examination of grey prediction model GM (1, 1)

Goodness of Fit Test showed that prediction accuracy grade (C) is 0.40; small error probability (P) is .91; the grade of prediction accuracy was 2 (qualified). So the established model can be used in extrapolated prediction. Incidences of TPF from 2004 to 2015 in Wuhan city were fitted and the results showed that estimated values coincided with the actual values. Actual values in 2016 was 0.22/100,000, estimated values in 2016 was 0.08/100,000.

The average of absolute residual value was 0.10/100,000 (Table 2).

3.4. Prediction of TPF incidence

Based on the established functional equation, a short-term extrapolated prediction was carried out to estimate incidences of TPF from 2017 to 2019. The results were listed in Table 2. Extrapolated prediction showed that the incidences of TPF from 2017 to 2019 were 0.08/100,000, 0.07/100,000, and 0.05/100,000, respectively. The incidences of TPF were obviously slowly decreasing.

4. Discussion

Time series prediction refers to the process by which the future values of a system is forecasted based on the information obtained from the past and current data points.^[25] Generally, a predefined mathematical model is used to make accurate predictions. Statistical and artificial intelligence-based approaches are the 2 main techniques for time series prediction seen in the literature. However, these techniques are not accurate for nonlinear problems. More importantly, they need large number of samples and are too complex to be used in predicting future values.^[26]

In contrast, grey system theory is a better alternative method for time series prediction, which is designed to work with system

Table 2
Comparison of actual value and estimate value of typhoid and paratyphoid fevers incidence.

Year	Actual value (1/100 thousand)	Estimated value (1/100 thousand)	Residual error (1/100 thousand)	Relative error (1/100)
2004	1.17	—	—	—
2005	1.03	1.01	0.02	1.94
2006	0.88	0.81	0.07	7.95
2007	0.52	0.66	-0.14	-26.92
2008	0.36	0.54	-0.18	-50.00
2009	0.67	0.43	0.24	35.82
2010	0.43	0.35	0.08	18.60
2011	0.16	0.29	-0.13	-81.25
2012	0.1	0.23	-0.13	-130.00
2013	0.24	0.18	0.06	25.00
2014	0.21	0.16	0.05	23.81
2015	0.13	0.12	0.01	7.69
2016	0.22	0.10	0.12	54.55
2017	—	0.08	—	—
2018	—	0.07	—	—
2019	—	0.05	—	—

in which the available information is insufficient to characterize the system.^[27] In systems theory, a system can be defined with a color that represents the amount of clear information about that system. For instance, if the information is known entirely, the system is called a white system. If the information is unknown, it is called a black system. If the information is being incomplete, it can be named as a grey system. Strictly, every system can be considered as a grey system because there are always some uncertainties. Because of the noise from both inside and outside of the system, the information we can reach about that system is always uncertain and limited in scope.^[21,27]

As superiority to conventional statistical approaches, grey system theory requires only a limited number of data to estimate the behavior of unknown systems, which is different from the previous methods. Because of its simple calculation process and higher forecasting accuracy, grey system theory has been widely used in the prediction of a lot of fields. In recent years, grey system theory has become more and more popular in biomedical information and technology. In grey system theory, GM (1, 1) model is an effective approach, which can make use of relatively small data sets and does not require to comply with certain statistical laws strictly, simple or linear relationships among the observable variables. Thus, it can overcome the disadvantages of statistical method.^[24] Therefore, this study adopted GM (1, 1) model to predict the incidence of TPF in Wuhan City and provided the results as reference for future studies and policy makers.

In the present study, based on the raw data of TPF incidence from 2004 to 2015 in Wuhan City, GM (1, 1) model was built to forecast the incidence in the next 3 years. The model accuracy examination results show that GM (1, 1) model is able to make accurate predictions for forecasting incidence of TPF. We compared actual value with predicted value in 2016. The result showed that predicted values are consistent with actual values, indicating that GM (1, 1) model we built is credible and effective in practice. Traditional TPF incidences estimation methods use statistics analysis, so that large data samples are required. With restricted conditions, it usually causes the results lack of authenticity and unsuitable to apply in practical use. By using the grey system modeling, the more reliable prediction can be obtained for future policy making in TPF prevention. In addition, incidences of infectious diseases are deeply influenced by social and natural factors, so database of grey model should be updated in time for long-term analysis.

According to the prediction of the TPF incidences, we found that the TPF incidences in Wuhan City will be slowly decreasing in the future 3 years. However, we should know that TPF still causes approximately 200,000 deaths annually and >90% of TPF cases are estimated to occur in Asia, owing to the consumption of unsafe drinking water, inadequate sewage disposal, and flooding.^[28] The recent increase in fluoroquinolone resistance of *S enterica* serotype Typhi has raised concerns due to the limited treatment options available in TPF endemic countries.^[29–31] How to prevent TPF effectively is still a big challenge in less-industrialized countries, for preventive measures are vital to reduce the occurrence of typhoid fever and avoid new outbreaks and effective prevention will result in large cost savings to the national health care system.

Contaminated water and food are important vehicles for transmission of TPF and preventive public health measures based on sanitation and hygiene have proved to be essential to the reduction of TPF. Therefore, careful food preparation and washing of hands are crucial in preventing TPF. Adequate water

treatment, waste disposal, and protection of the food supply from contamination are also important public health measures. In addition, carriers of TPF must not be allowed to work as food handlers. And finally, the use of TPF vaccines will be helpful to reduce the susceptibility of hosts to infection.

It was reported that the incidence of TPF was high among the residents of the densely populated urban community.^[32] As low-income workers are increasingly attracted to urban centers and rural-urban fringe zone with available jobs, the population residing in informal settlements will not have access to the available water and sanitary infrastructure. The incidence of TPF demonstrates the need for longer-term investment in improvement of water and sanitation services to reduce the burden of multiple fecal-oral transmitted pathogens in these communities. Targeted vaccination against TPF would be a valuable immediate step to reduce disease burden, especially in densely populated urban community. However, more emphasis should be placed upon sanitary improvements and health education, rather than focusing solely on improving the health delivery system.

It is also likely that an important proportion of cases are due to travels. Pretravel screening and vaccination strategies are essential measures for travelers to endemic areas.

Our present study must be interpreted in light of some limitations. Firstly, TPF cases reporting guideline were “GB 16001-1995” (2004–2007) and “WS 280-2008” (2008–2015), respectively. There are 3 slight differences between these 2 versions [(a) field epidemiological survey criterion in “WS 280-2008” not only contains factors such as epidemic area in “GB 16001-1995,” but also consists of specific time, close contact history, and hygienic habits. (b) unexplained persistent high fever was defined as clinical symptom in “WS 280-2008,” which was more specific than persistent high fever in “GB 16001-1995”. (c) in “WS 280-2008,” laboratory test results could present the count of eosinophilic granulocytes were decreased or disappeared, and the total number of white blood cells were normal or lower, whereas in “GB 16001-1995,” it described as the count of eosinophilic granulocytes disappeared, and the total number of white blood cells were lower], and it seemed the criterion of “WS 280-2008” is more extensive and more precise, which may slightly influence the efficacy of prediction. Secondly, we have not construct train set, because the data are actually not big enough for training a GM model if being divided into train set, validation set, and test set. Therefore, expanding the data of TPF should be considered for future research.

In conclusion, the purpose of the present study was to adopt grey system theory to predict the incidence trend of TPF in Wuhan City, and provide reference for the government in policy making. Our study shows that grey forecasting model GM (1, 1) we built is an effective forecast method and can be used to predict the incidence of TPF. This finding may serve as a reference to future studies and policy making.

Acknowledgments

This work was supported by the Young and Middle-aged Medical Scholar Training Project of Wuhan from *Health and Family Planning Commission of Wuhan Municipality*.

Author contributions

Conceived and designed the experiments: GFJ. Analyzed the data: JJZ. Contributed reagents/materials/analysis tools: DGK. Wrote the article: XBY.

Conceptualization: Gaofeng Jiang.

Formal analysis: Jiaojiao Zou.

Resources: Deguang Kong.

Writing – original draft: Xiaobing Yang.

Writing – review and editing: Xiaobing Yang.

Author name: ORCID: 0000-0002-2472-5730

References

- [1] Wang LX, Li XJ, Fang LQ, et al. Association between the incidence of typhoid and paratyphoid fever and meteorological variables in Guizhou, China. *Chin Med J (Engl)* 2012;125:455–60.
- [2] Amicizia D, Arata L, Zangrillo F, et al. Overview of the impact of typhoid and paratyphoid fever. Utility of Ty21a vaccine (Vivotif (R)). *J Prev Med Hyg* 2017;58:1–8.
- [3] Cutler D, Miller G. The role of public health improvements in health advances: the twentieth-century United States. *Demography* 2005;42:1–22.
- [4] Obaro SK, Iroh Tam PY, Mintz ED. The unrecognized burden of typhoid fever. *Expert Rev Vaccines* 2017;16:249–60.
- [5] Liu FF, Zhao SL, Chen Q, et al. Surveillance data on typhoid fever and paratyphoid fever in 2015, China. *Zhonghua Liu Xing Bing Xue Za Zhi* 2017;38:754–8.
- [6] Wang JF, Wang Y, Zhang J, et al. Spatiotemporal transmission and determinants of typhoid and paratyphoid fever in Hongta District, Yunnan Province, China. *PLoS Negl Trop Dis* 2013;7:e2112.
- [7] Gu H, Fan W, Liu K, et al. Spatio-temporal variations of typhoid and paratyphoid fevers in Zhejiang Province, China from 2005 to 2015. *Sci Rep* 2017;7:5780.
- [8] Sun JL, Zhang J, Ma HL, et al. Epidemiological features of typhoid/paratyphoid fever in provinces with high incidence rate and in the whole country, in 2012. *Zhonghua Liu Xing Bing Xue Za Zhi* 2013;34:1183–8.
- [9] Vollaard AM, Ali S, Van Asten HA, et al. Risk factors for typhoid and paratyphoid fever in Jakarta, Indonesia. *JAMA* 2004;291:2607–15.
- [10] Bhan MK, Bahl R, Bhatnagar S. Typhoid and paratyphoid fever. *Lancet* 2005;366:749–62.
- [11] Sun L, Shao Q, Wang ZQ, et al. Spatial structure of rodent populations and infection patterns of hantavirus in seven villages of Shandong Province from February 2006 to January 2007. *Chin Med J (Engl)* 2011;124:1639–46.
- [12] Kelly-Hope LA, Alonso WJ, Thiem VD, et al. Temporal trends and climatic factors associated with bacterial enteric diseases in Vietnam, 1991–2001. *Environ Health Perspect* 2008;116:7–12.
- [13] McIver LJ, Imai C, Buettner PG, et al. Diarrheal diseases and climate change in Cambodia. *Asia Pac J Public Health* 2016;28:576–85.
- [14] McMichael AJ, Woodruff RE, Hales S. Climate change and human health: present and future risks. *Lancet* 2006;367:859–69.
- [15] Zhang Y, Bi P, Hiller JE, et al. Climate variations and bacillary dysentery in northern and southern cities of China. *J Infect* 2007;55:194–200.
- [16] Shuman EK. Global climate change and infectious diseases. *N Engl J Med* 2010;362:1061–3.
- [17] Deng JL. Introduction to grey system theory. *J Grey System* 1989;1:1–24.
- [18] Bindhu BK, Madhu G. Application of grey system theory on the influencing parameters of aerobic granulation in SBR. *Environ Technol* 2017;38:2143–52.
- [19] Qiu WR, Sun BQ, Xiao X, et al. iPhos-PseEvo: identifying human phosphorylated proteins by incorporating evolutionary information into general PseAAC via grey system theory. *Mol Inform* 2017;36:
- [20] Wang CN, Nguyen NT, Tran TT. Integrated DEA models and grey system theory to evaluate past-to-future performance: a case of Indian electricity industry. *Scientific WorldJournal* 2015;2015:638710.
- [21] Zhang C, Zhang H. Analysis of aerobic granular sludge formation based on grey system theory. *J Environ Sci (China)* 2013;25:710–6.
- [22] Xiao X, Hui MJ, Liu Z, et al. iCataly-PseAAC: identification of enzymes catalytic sites using sequence evolution information with grey model GM (2,1). *J Membr Biol* 2015;248:1033–41.
- [23] Mahmod WE, Watanabe K. Modified Grey Model and its application to groundwater flow analysis with limited hydrogeological data: a case study of the Nubian Sandstone, Kharga Oasis, Egypt. *Environ Monit Assess* 2014;186:1063–81.
- [24] Bao CZ, Mayila M, Ye ZH, et al. Forecasting and analyzing the disease burden of aged population in china, based on the 2010 global burden of disease study. *Int J Environ Res Public Health* 2015;12:7172–84.
- [25] Gan R, Chen X, Yan Y, et al. Application of a hybrid method combining grey model and back propagation artificial neural networks to forecast hepatitis B in china. *Comput Math Methods Med* 2015;2015:328273.
- [26] Amigo JM, Hirata Y, Aihara K. On the limits of probabilistic forecasting in nonlinear time series analysis II: differential entropy. *Chaos* 2017;27:083125.
- [27] Huang JC. Application of grey system theory in telecare. *Comput Biol Med* 2011;41:302–6.
- [28] Crump JA, Luby SP, Mintz ED. The global burden of typhoid fever. *Bull World Health Organ* 2004;82:346–53.
- [29] Chau TT, Campbell JI, Galindo CM, et al. Antimicrobial drug resistance of *Salmonella enterica* serovar typhi in Asia and molecular mechanism of reduced susceptibility to the fluoroquinolones. *Antimicrob Agents Chemother* 2007;51:4315–23.
- [30] Parry CM. The treatment of multidrug-resistant and nalidixic acid-resistant typhoid fever in Vietnam. *Trans R Soc Trop Med Hyg* 2004;98:413–22.
- [31] Slinger R, Desjardins M, McCarthy AE, et al. Suboptimal clinical response to ciprofloxacin in patients with enteric fever due to *Salmonella* spp. with reduced fluoroquinolone susceptibility: a case series. *BMC Infect Dis* 2004;4:36.
- [32] Naheed A, Ram PK, Brooks WA, et al. Burden of typhoid and paratyphoid fever in a densely populated urban community, Dhaka, Bangladesh. *Int J Infect Dis* 2010;14(suppl 3):e93–9.