

PREDICTION MARKET SYSTEM FOR EPIDEMIC DISEASE PREDICTION

Eldon Y. Li, National Chengchi University, 64, Sec. 2, Zhi-nan Rd., Wenshan Taipei 11605, Taiwan, eli@calpoly.edu, +886-2-2939-3091 ext. 81203

Chen-Yuan Tung, National Chengchi University, 64, Sec. 2, Zhi-nan Rd., Wenshan Taipei 11605, Taiwan, ctung@nccu.edu.tw, +886-2-2938-7286

Shu-Hsun Chang, National Chengchi University, 64, Sec. 2, Zhi-nan Rd., Wenshan Taipei 11605, Taiwan, 98356504@nccu.edu.tw, +886-2-2939-3091 ext. 85006

ABSTRACT

The quest for an effective system capable of monitoring and predicting the trends of epidemic disease is a critical issue nowadays. This study designed and constructed an Epidemic Prediction Market System (EPMS) with Market Scoring Rules (MSR) matching mechanism. The system was used to predict infectious diseases with collective wisdom from 126 medical professionals during the period of thirty one weeks in 2010. The results showed that this system with MSR mechanism is more accurate in predicting epidemic disease trends than the system using expected value of historical data for the same period. Furthermore, the EPMS can continuously update data to improve the performance of prediction. It is highly suitable for applying to a wider range of diseases and geographical areas, especially for under-developed areas in Asia and Africa.

Keywords: Prediction Market, Market Scoring Rules, Epidemic Prediction

INTRODUCTION

Epidemic diseases cause thousands of people to lose their lives each year. According to the World Health Organization (WHO), annual cases of cholera, one of the most common epidemic diseases, range from 1.4 to 4.3 million, causing 28 to 142 thousand deaths each year (WHO, 2012). This number holds despite the well-established and effectively-executed cholera prevention and treatment programs in many regions of the world. The quest for an effective system capable of monitoring and predicting the trends of epidemic diseases is a critical issue for worldwide communities.

Recently, Google introduced “Flu Trends” application (<http://www.google.org/flutrends/>) to monitor the severity of flu epidemic in most developed countries by using aggregated Google search data to estimate current flu activity around the world in near real-time (Ginsberg et al.,

2009). Polgreen et al. (2008) used a similar tool, Yahoo! search engine, to produce flu trends that are strikingly similar to actual influenza occurrence pattern. Although both studies advocate the use of search engine to predict influenza epidemic, there are three limitations of Google's flu trend application. First of all, web searches for real-time influenza data can indicate what had happened, but not what will happen; i.e., they are not predictions if we define the term rigorously. Besides, Google Flu Trends does not provide surveillance signals for most regions in Asia and Africa. Moreover, this application can only be used to monitor the epidemic of flu. It is not applicable in the cases of other diseases. It seems that currently there is no suitable and reliable tool to predict the trends of epidemic diseases.

One plausible alternative to overcome the aforesaid difficulty in having real-time prediction is to use an epidemic prediction market system (EPMS). Prediction markets are useful for forecasting the spread of infectious diseases (Polgreen et al., 2006; Polgreen et al., 2007). They can continuously renew the data to make an up-to-date prediction. The purpose of this study is to demonstrate the effectiveness of EPMS for predicting the trends of epidemic diseases. We first constructed an EPMS specifically for EDP and invited thousands of medical professionals in major hospitals and health institutions in Taiwan to use the system. The results confirm high accuracy of EDP provided by the EPMS. Both academicians and practitioners could apply the EPMS of this study to create their own systems for EDP in their own regions.

PREDICTION MARKET SYSTEM

Prediction market system (PMS), operating like a futures market, can be used as a mechanism to integrate information from different sources to predict the outcomes of future events. PMS have two major features: provide appropriate incentives and punishment mechanisms and perform continuous corrections. Traders in a PMS make transactions for the predictions of a future event based on public or private information. The actual results of the future event determine the rewards (or penalties) for the trader, whereas the price of the future event contract represents the entire market's prediction of the outcome of the future event.

Participants in PMS can continuously express their opinions during the certain time period, and they can also get the fairer return proportionally. Indeed, prediction markets have a high level of prediction accuracy. This paper uses a different matching rule of prediction markets to improve the effectiveness of forecasting the epidemics of infectious diseases in Taiwan.

Infectious diseases in Taiwan are difficult to predict and is thus a good case to test the capabilities of prediction markets. Located in the subtropical climate region, Taiwan is subject to easy trends of infectious diseases. Global warming and international transportation further increase the frequency and seriousness of these diseases, hence the difficulty to predict them. The EPMS we constructed not only provided accurate and real-time predictions as other prediction markets did for other diseases, but also exceeded the Iowa Influenza Market (IIM), so far the most successful prediction market of epidemics, in terms of duration, space, the number of diseases, and the method of transaction.

SYSTEM EXPERIMENT AND RESULTS

Sample Screen

FIGURE 1 shows the view of prediction event list in the EPMS. Participant could choose the prediction area from the Pull-down menus in the upper right corner of the view. When choosing a specific prediction event by clicking the event name, the system would navigate the participant to trading view page. Participant could make a transaction after viewing the orders. In the view page, participant needed to set the prediction range and place the bets for each prediction event. Then, the EPMS would automatically calculate his or her expected return of each bet. Based on the prediction result, we compared the results of EPMS with the historical data.

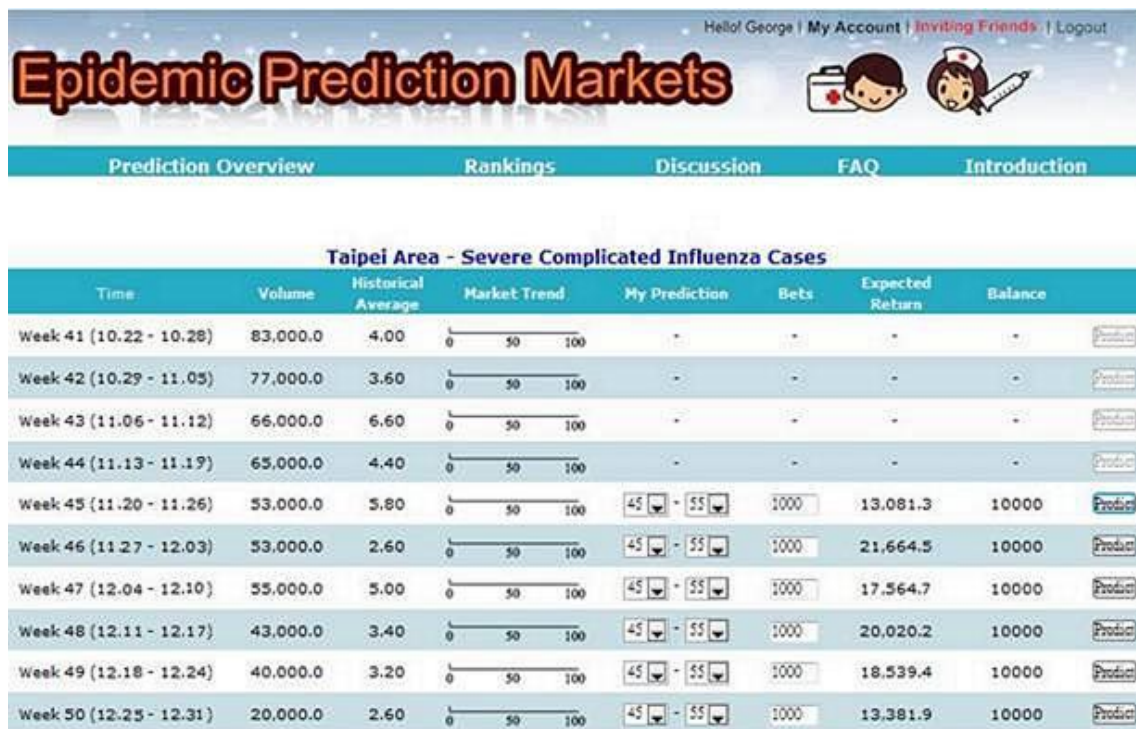


FIGURE 1: The Trading View of EPMS

Participation of EPMS

From the 10th week (March 7-13) to the 40th week (October 3-9) of 2010, 630 medical professionals registered with this EPMS. However, only 126 members traded the prediction events of diseases indicators. The composition of these participants were: 48 nurse specialists, 23 doctors, 13 medical inspectors, four pharmacists, three nurses, two Chinese medicine practitioners and 33 other professionals. Regarding working institutions, 40 participants were from public hospitals and another 40 from private hospitals; 21 from Taiwan's Ministry of Health and its related units; 7 from clinics; 4 from local public health bureaus; the rest 14 participants from other units. In terms of areas where the participants were working, 38 participants came from Taipei area; 29 from southern area; 21 from northern area; 15 from Kaohsiung-Pingtung area in the south of Taiwan; 12 from eastern area; and 11 from central area. TABLE 1 shows the statistics of participants participating in the prediction of five diseases indicators.

TABLE 1: The Statistics of Participants Participating in the Prediction of Five Diseases Indicators

Indicator	Participants	Predictions	Trading amount
confirmed cases of dengue fever	84	4,395	29,503,332
confirmed cases of severe complicated influenza case	81	3,956	33,501,580
rate of enterovirus infection	76	3,495	31,627,923
rate of influenza-like illness	64	2,660	23,759,804
confirmed cases of severe complicated enterovirus infection	59	4,291	23,199,675
Total	126	18,797	141,592,314

Forecasting Accuracy

In this paper, we compared the forecasting accuracy of two forecasting methods: the first is the historical average number of disease cases or rate (AVG) for the same weeks from 2005 to 2009, which is adopted by the Centers for Diseases Control (CDC) in Taiwan for the prediction of epidemic diseases. The second is the predictive value of the EPMS. In order to compare the accuracy, we calculated the prediction error values of these two methods as follows:

$$\text{Prediction_error_value} = \text{Predicted_value} - \text{Actual_value}$$

Lower prediction error value means higher prediction accuracy, and vice versa. When the prediction error value of EPMS is less than that of AVG, the number of wins for EPMS is to increase by one. Therefore, the winning ratio is calculated as follows:

$$\text{Winning_ratio} = \text{Wins} / \text{Total_number_of_predictions}$$

Winning Ratios of Two Methods for All Diseases and All Areas

If every week's prediction on each diseases indicator was regarded as a prediction event, there were 7,945 prediction events in total. Concerning the prediction performance, for the target week (0 week in advance), EPMS was more accurate in 701 out of 1,085 prediction events than AVG and the winning ratio of EPMS versus AVG was 64.6% (see FIGURE 2). EPMS's winning ratio was 55.5% for 1 week in advance, 54.4% for 2 weeks in advance, 53.0% for 3 weeks in advance, 52.8% for 4 weeks in advance, 52.3% for 5 weeks in advance and 50.5% for 6 weeks in advance. The winning ratio of EPMS was only inferior to that of AVG for the 7 weeks in advance.

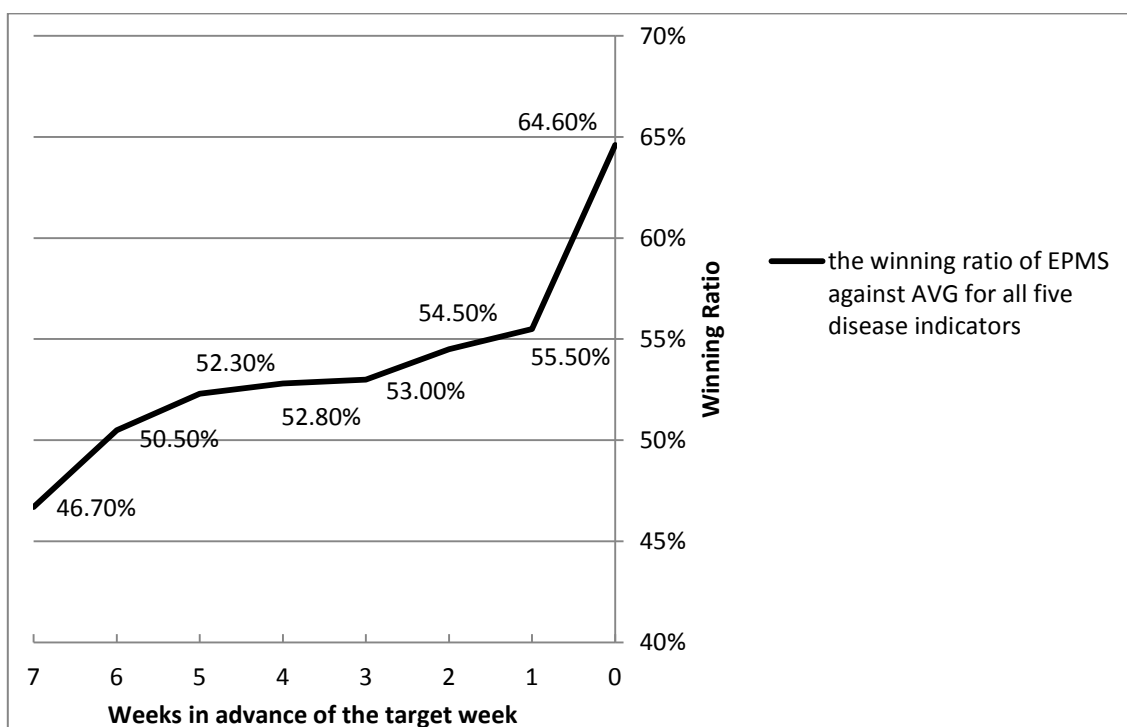


FIGURE 2: The Winning Ratio of EPMS against AVG for All Five Indicators

Prediction Errors and Winning Ratios of Two Methods for Each Disease and All Areas

For the absolute prediction error of all five diseases indicators, EPMS was more accurate than AVG except for the dengue fever confirmed cases (See TABLE 2). For the prediction of the confirmed cases of severe complicated influenza, EPMS's prediction error was 11.955 cases while AVG's prediction error was 18.206 cases. For the prediction of the confirmed cases of dengue fever, EPMS's prediction error was 18.131 cases while AVG's prediction error was 16.509 cases. For the prediction of the confirmed cases of severe complicated enterovirus infection, EPMS's prediction error was 1.172 cases while AVG's prediction error was 2.935 cases. For the prediction of the rate of enterovirus infection, EPMS's prediction error was 0.0184% while AVG's prediction error was 0.0411%. For the prediction of the rate of influenza-like illness, EPMS's prediction error was 0.0104% while AVG's prediction error was 0.0132%.

TABLE 2: Prediction Error of EPMS and AVG for Five Diseases Indicators

Indicator	EPMS	AVG
Confirmed cases of severe complicated influenza case	11.955	18.206
Confirmed cases of dengue fever	18.131	16.509
Confirmed cases of severe complicated enterovirus infection	1.172	2.935
Rate of enterovirus infection (%)	0.0184	0.0411
Rate of influenza-like illness (%)	0.0104	0.0132

Cross Analysis of Prediction of Five Diseases and Eight Weeks

For each disease indicator, there were 1,589 prediction events, including 217 prediction events for both target week and 1 week in advance of the target week, 210 for 2 weeks in advance, 203 for 3 weeks in advance, 196 for 4 weeks in advance, 189 for 5 weeks in advance, 182 for 6 weeks in advance, 175 for 7 weeks in advance. For the prediction for the confirmed cases of severe complicated influenza, the rate of enterovirus infection, and the rate of influenza-like illness, the winning ratios of EPMS versus AVG were clearly over 50% in all weeks. The winning ratios of EPMS for these three indicators in the target week were 69.6%, 83.9% and 76.0%, respectively. Instead, for the prediction of the confirmed cases of dengue fever and the confirmed cases of severe complicated enterovirus infection, the winning ratios of EPMS versus AVG were all below 50%. Particularly, the winning ratios of EPMS for the prediction of the confirmed cases of severe complicated enterovirus infection were all under 31% prior to the target week.

Participation of EPMS

From the 10th week (March 7-13) to the 40th week (October 3-9) of 2010, 630 medical professionals registered with this EPMS. However, only 126 members traded the prediction events of diseases indicators. The composition of these participants were: 48 nurse specialists, 23 doctors, 13 medical inspectors, four pharmacists, three nurses, two Chinese medicine practitioners and 33 other professionals. Regarding working institutions, 40 participants were from public hospitals and another 40 from private hospitals; 21 from Taiwan's Ministry of Health and its related units; 7 from clinics; 4 from local public health bureaus; the rest 14 participants from other units. In terms of areas where the participants were working, 38 participants came from Taipei area; 29 from southern area; 21 from northern area; 15 from Kaohsiung-Pingtung area in the south of Taiwan; 12 from eastern area; and 11 from central area.

CONCLUSION

This study constructed an EPMS with MSR matching mechanism. The system was used to predict infectious diseases with 126 medical professionals during the period of thirty one weeks in 2010. The prediction results from the EPMS showed that winning ratio of the EPMS for epidemic predictions was constantly increasing along with approaching target weeks of expiry date. Due to the closed environment of professionals, this EPMS with MSR mechanism proved to be more accurate in predicting epidemic disease trends comparing with the expected value of historical data for the same period. Indeed, extant approaches can only be used to monitor the condition of diseases, not to make predictions. EPMS can continuously update data to correctly predict the trends of diseases.

In addition, the EPMS is highly suitable for application in cases of a wider range of diseases and geographical areas vis-à-vis Google Flu Trends, which overlooked most of the less developed regions in Asia and Africa. These regions are known for being prone to contagion of infectious diseases. Google Flu Trends can only be used for monitoring flu trends, while the EPMS could be applied to predicting, detecting, and monitoring the trends of infectious diseases. Overall, the EPMS is an effective instrument to deal with the epidemic management and prevention in practice for the CDC.

REFERENCES

1. Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature* 457(7232), 1012-1015.
2. Polgreen, P. M., Chen, Y., David, M., Pennock, D. M., & Nelson, F. D. (2008). Using internet searches for influenza surveillance. *Clinical Infectious Diseases* 47(11), 1443-1448.
3. Polgreen, P. M., Nelson, F.D., & Neumann, G. R. (2007). Use of prediction markets to forecast infectious disease activity. *Clinical Infectious Diseases* 44(2), 272-279.
4. Polgreen, P. M., Nelson, F. D., & Neumann, G. R. (2006). Using prediction markets to forecast trends in infectious diseases. *Microbe* 1(10), 459-465.
5. World Health Organization (2012). WHO Technical Working Group on creation of an oral cholera vaccine stockpile, Meeting report, Geneva, 26-27 April.