Modeling Disease Spread at Global Mass Gatherings: Data Requirements and Challenges

Sultanah M. Alshammari and Armin R. Mikler

Abstract Spread of infectious diseases at global mass gatherings can pose health threats to both the hosting country and the countries where participants originate. The travel patterns at the end of these international events may result in epidemics that can grow to pandemic levels within a short period of time. Computational models are essential tools to estimate, study, and control disease outbreaks at mass gatherings. These models can be integrated in the planning and preparation process of mass gatherings. In this paper, we present a review of the key data requirements and the challenges encountered when modeling infectious diseases epidemics initiated by global mass gatherings. This review can assist epidemic modeling at global mass gatherings providing researchers with insights of the main aspects and possible big data opportunities in this emerging research area.

Keywords Mass gatherings (MGs) · Disease spread · Infectious diseases · Epidemic · Outbreak · Computational modeling

1 Introduction

There are several recurrent global mass gatherings (MGs) occurring on different places in the world such as sport, religious, and political events. Participants attending these events may be exposed to serious health threats. The gathering of high population density in a close proximity during MGs has the potential to facilitate the transmission of communicable diseases [1–5]. Furthermore, the global

S.M. Alshammari (✉)
Department of Computer Science & Engineering, University of North Texas,
Denton, TX, USA
e-mail: SultanahAlshammari@my.unt.edu

A.R. Mikler
Center for Computational Epidemiology and Response Analysis (CeCERA),
University of North Texas, Denton, TX, USA
e-mail: Armin.Mikler@unt.edu

events (such as the Olympics, the FIFA World Cup, and the Hajj; the Muslim pilgrimage to Makkah) where people gather from different countries present serious health threats and challenges for both the hosting countries and the countries of origin of the participants. These global MGs allow the mixing of various infectious pathogens due to the diverse disease exposure history and demographics of the participants. When these global events are over, the travel patterns of the participants could cause a rapid spread of infectious diseases affecting large number of people and causing global epidemic within a short period.

There was a rising concern in 2009, as MGs occurring that year such as the Hajj and the Southeast Asian Games could further contribute to global spread of the 2009 pandemic influenza A (H1N1) [6]. In fact, after the 2009 Hajj season several studies [7–10] were conducted to trace H1N1 pandemics among the returning pilgrims and their contacts. Also, activities of the 2009 pandemic influenza A (H1N1) virus have been reported during several music festivals [11]. Several outbreaks have been reported in different MGs of varying sizes with the most common reported infections being respiratory viruses including different strains of influenza [12]. As shown in the Fig. 1 below, there was a global spread of N meningitidis W135 after recurrent outbreaks of W135 during the Hajj seasons of 2000 and 2001. Also, influenza outbreaks have been reported in 2002 and in 2008 at the Winter Olympics in Salt Lake City, UT, USA [13], and the World Youth Day in Sydney [14] respectively. In 2008, at the European Football Championship several measles outbreaks occurred in both hosting countries Austria and
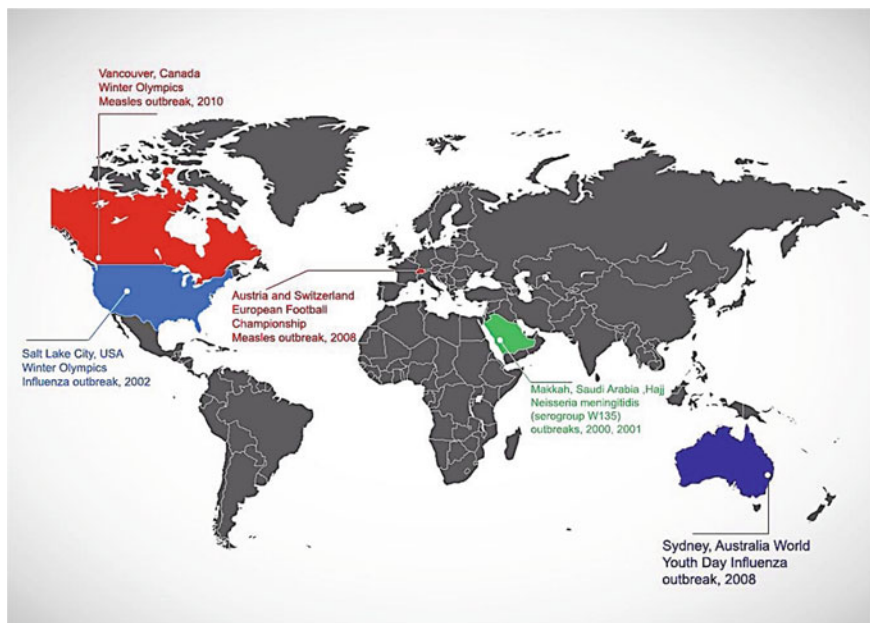

Fig. 1 Some reported outbreaks in previous mass gatherings

Switzerland, with other measles cases reported in the neighboring countries of France, Germany, and Spain [15].

There are increasing research efforts addressing the health hazards in the context of mass gatherings and their impacts on global health. Many studies [2, 5, 16–19] draw attention to major threats presented in mass gatherings, and provide guidelines and recommendations to manage possible health hazards. In their study, Nsoesie et al. [20] reviewed the use of new technologies to improve infectious diseases surveillance in different types of MGs as a part of planning and preparation for these events. These technologies include web-based systems, smartphone applications, wireless sensor networks and syndromic surveillance systems. While, applying all the different measures and efforts can be effective in preventing and controlling the spread of infectious diseases during mass gatherings, it may not be sufficient to provide insights of the potential risk of global epidemics after these events. Computational methods and models can play an important role not only in the prevention and control of disease outbreaks in mass gatherings, but also in predicting a global spread of infectious diseases.

In this paper, we aim to summaries the main aspects of modeling disease spread at mass gatherings, and provide a review over relevant literature. Based on this review, the key data requirements to model disease spread at global MGs and the encountered challenges related to big data are presented. The paper is organized as follows; the next section summaries several recent computational models of infectious disease in mass gatherings. Next, data requirements and challenges in modeling epidemics at global MGs are covered. Then, some big data aspects and opportunities when modeling disease spread at MGs are highlighted. Finally, the last section summaries the review presented in this paper.

2   Modeling Disease Outbreaks at MGs

The success of computational modeling in epidemiology and public health motivates applying its use in the context of mass gatherings to provide efficient tools to study the spread of diseases at mass gatherings and assess the risk of global epidemics. There are several studies proposing computational models for analyzing and simulating epidemics of different diseases in different settings and populations. These models vary from the simple mathematical SIR [21] (Susceptible--Infected–Recovered) model to more complicated spatial-temporal simulations that allow the introduction of prevention and mitigation measures. However, few studies are devoted to computational models of epidemics at mass gatherings. In their study, Chowell et al. [22] examined different aspects of epidemics that must be addressed in computational models including; the environmental, behavioral, health-related, and demographics factors and highlighted several data requirements and modeling challenges in these settings. Their article concluded with recommendations of integrating disease spread models as a part of the preparedness and the planning of some of the global events such as the Olympics and the Hajj.

In an event-based study, Khan et al. [23] proposed a conceptual model to integrate data of international air travels from different sources and reported cases of infectious diseases as a preparation for potential infectious diseases threats at the Vancouver 2010 winter Olympics. This work suggested that the hosting country of such global event could use this integrated knowledge model to estimate emerging health threats. First, the authors gathered and analyzed the previous global travel patterns, focusing on the Olympic Winter Games for the past decade. Then, based on results of that analysis, the 25 cities with the highest numbers of travellers were identified, and the infectious disease surveillance in those cities was conducted. Next, the information from the HealthMap (www.healthmap.org) was integrated to the global infectious disease surveillance. As described by Khan et al. [23], HealthMap is an online tool that tracks outbreaks reports from different sources and transforms these reports into interactive maps. HealthMap provided a real time analysis before and during the 2010 Olympic Winter Games. Although, applying their model to the 2010 Olympic Winter Games did not reveal any serious threats, it provided the Canadian health officials with advanced knowledge to be prepared to any threats.

While Khan et al. [23] targeted one of the global mass gatherings to identify possible incoming health hazards, Stehle et al. [24] proposed a simulation of infectious disease transmission for a relatively small gathering among the attendees of the 2009 Annual French Conference on Nosocomial Infections. They used a stochastic SEIR (Susceptible–Exposed-Infected–Recovered) model to demonstrate the impact of using different contact patterns on the dynamic of the infection. The authors were able to measure the interactions between the individuals within the two days of the conference using wearable sensors (RFID; radiofrequency identification devices). Then, the collected data was used to compare three different contact networks: dynamic, homogeneous, and heterogeneous. These contact networks differ in the type and the duration of the interactions between individuals. Their results of the contact duration and distribution showed a high probability of pathogen spread for a small number of infected people. This probability was smaller in homogeneous networks as compared to heterogeneous and dynamic networks.

Shi et al. [25] developed a SEIR agent-based simulation model using data from the state of Georgia to study the impacts of the contact patterns and social mixing on influenza pandemic in different settings including mass gatherings and holiday traveling. In their simulation, they divided the simulation period into a "regular" period and a "traveling" or "mass gathering" period. In the "regular" period, Agents are moving to and from households, schools or workplaces, and other public places. The proportion of the created population was selected to mix within a large group to model temporal mass gatherings. To validate their model, Shi et al. [25] executed the simulation using several scenarios of different populations interacting in mass gatherings or traveling periods varying in length and size. The results obtained from the simulation experiments identified several factors that affect the dynamics of the influenza epidemic, including the timing of mass gatherings or travel in relation to the epidemic peak and the pathogen's infectious period. Also, the simulation indicated little impact of the changes in the social mixing on the course of an

epidemic. The authors suggested that using such predictive model could provide public health officials with insights to take proper actions and control measures during a mass gathering.

3 Data Requirements

As stated by Chowell et al. [22], the risk of disease spread at MGs is related to the participants attending these events and the environment where the event took place. Thus, the key data requirements to model outbreaks in MGs include the characteristics of the participating populations and the event itself.

3.1 Participating Populations

The main attributes of the population attending MGs that should be used to guide modeling disease spread, include the population demographic and epidemiological disease-specific aspects such as susceptibility levels and vaccination status. The most important demographic data about the participants attending a global event are *age*, *gender*, and the *country of origin*. In fact, based on the epidemic modeling approach these demographic data if known, can be used to identify, estimate, or make reliable assumptions about other modeling attributes such as the contact patterns, the susceptibility levels, and the vaccination status. Knowing the age distribution of the population can provide insights into the contact rates and the susceptibility levels of different age groups. In fact, the correlation between the different age groups and the contact rates was investigated in several studies. In a recent study [26], Cowling et al. revealed a strong age-based contact rate based on analyzing contact patterns in different settings in several European countries. It was concluded that compared to other age groups, children and adolescents tend to have a higher rate of contacts within their age group. Thus, children and teenagers play an important role in disease spread of close contact infections especially influenza. As a result it is important to conduct studies for age-based risk groups for different diseases in the context of MGs.

Gender is another important factor when modeling disease transmission among individuals. Gender differences can contribute in understanding and controlling the transmission of infectious diseases [27]. However, most studies investigated gender differences on emerging infectious diseases focused mainly on sexual transmitted diseases. During mass gatherings, the gender distribution of the population is important to study and model disease spread especially when combined with gender-based behavior analysis within the event. As in some MGs especially the religious events, there is a gender separation either for the entire event or in some stages of the event. For example, at the Hajj, men and women are only segregated in

their sleeping arrangements, as opposed to another annual religious event of Attukal Pongala festival in Kerala, India where only women are participating.

Countries of origins of the participants might provide insights of their susceptibility to some diseases, possible disease history, and vaccination status. As different countries have different vaccination policies and populations from different geographical areas vary in their disease exposure history. The incoming travel patterns can be used to determine the country of origin of the population during global MGs. In some MGs like Hajj, vaccination requirements for entry visa are issued based on the country of origin. For example, arrivals from polio-endemic countries are required to provide a proof of Poliomyelitis vaccination 6 weeks prior attending Hajj [28]. The vaccination coverage of different age groups can be acquired from different sources based on the country of origin such as official public health websites. The National Center for Health Statistics (NCHS) (www.cdc.gov/nchs/) provides detailed health measures and statistics including age-based vaccination coverage of different diseases. Also, there are several published reports and surveys provide information about the immunizations status in several countries either for specific age groups or the whole population. For example, the publicly available information provided in several reports [29, 30] of the seasonal influenza vaccination in Europe can be used to estimate the levels of immunity and susceptibility when modeling a MGs occurring in Europe.

3.2   Mass Gatherings Event

The different characteristics of the event such as timing, size, incoming travel patterns, setting, schedule of events, and the spatial considerations needed to be identified and included in the epidemic model. There is evidence of associations between climatic conditions and some infectious diseases such as seasonal influenza [7]. Thus, it is important to have a better understanding of climate contribution to the disease(s) under study and to integrate climate data into epidemic modeling. The expected weather patterns at MGs can be estimated based on the timing and the geographical location of the event. For example, in the past few years the Hajj season shifted from summer to winter [31]. As a result in the last seasons of Hajj, the summer related diseases such as heat stroke and food poisoning are not reported and the more expected diseases are related to cold weather such as influenza and asthma.

Arrival and departure are the most important phases of a global MGs, as the movement of large numbers of people from and to different countries to attend an event can pose public health threats to the hosting country. The incoming travel patterns can be used to determine the countries of origin of the participants, the expected diseases, and the possible vaccinations. Thus, identifying and analyzing the incoming travel patterns can help hosting countries to estimate the risks of importing infectious diseases. Moreover, when participants are travelling back to their countries after a global event this might contribute to a global epidemic within

a short period of time. For example, in their study Khan et al. [32] provided detailed analysis of the outgoing travel patterns from Mexico after the identification of the novel 2009 A (H1N1) influenza in Mexico and California. Their study confirmed H1N1 introductions within a few weeks in 20 countries where the highest number of arriving passengers from Mexico. Also, as stated previously after the Hajj season on 2009 several studies [7–10] were conducted to trace H1N1 pandemics among the returning pilgrims arriving from Saudi Arabia. A recent study by Khan et al. [33] analyzed 2012 international air travel data and Hajj data to predict pilgrims' movements after Hajj and estimate the potential novel coronavirus (MERS-CoV) importations.

Also, it is important to know and include the setting of the event; whether the event is held in one location or several locations; in a confined space or an open area, and whether it is a seated or mobile event. For example, in the Olympics Games and the Soccer World Cups several locations are involved. However, for most religious events, the rituals are performed at specific holy sites. Also, there is a need to determine the possible movements of the participants during the event; is there any specific regulations on their movements or their interactions with the local population within the hosting country. For instance, the movement of pilgrims attending Hajj is restricted and they are not allowed to extend their stay in the country after the completion of Hajj. All these aspects can play an important role in the dynamics of the disease spread and should be used to guide the modeling process of infectious disease in MGs.

4   Big Data and Modeling Disease Spread at MGs

There are several challenges related to the required data when attempting to model disease spread in global MGs. The variety of data to be captured from several sources is extensive, and includes demographics, airline traffic data, climate data, spatial data, and social and mobile phone network. Moreover, gathering detailed data about the expected participating populations at MGs prior to the event is very challenging if not infeasible. But, it is essential to have at least an adequate approximation of the demographics of the expected populations. For example, data from previous occurrences of MGs can be collected and analyzed to estimate the features of future participants attending similar events. In addition, as suggested by Nsoesie et al. [20] and Chowell et al. [22] using advanced monitoring and sensing devices of a representative sample provides higher levels of details about partici-pants at MGs and their interactions and movements. With such advanced methods extensive amounts of data will be generated and need to be analyzed to extract useful information about the participants, their demographics, contact patterns, and behaviors during the event. Social media such as twitter can provide an alternative data sources about these events. Analyzing tweets posted by the participants can provide insights about their behaviors, movements and the event itself, which can be used when modeling disease spread. Moreover, simulating vastly increasing and

high dimensional heterogeneous data representing the populations at global MGs provide a great big data challenge. As there is a need to integrate individual differences when modeling disease transmission at each stage of the event. In fact, some recent studies are aiming to include genetic variation among individuals when modeling infectious diseases and to use pathogens' genetic profiles to construct epidemics [33], which will result in having to process vast amount of genetic data.

Moreover, representing populations attending MGs in the epidemic model is a challenging task, as we want to capture differences on individual levels, we also need to represent specific subgroups among the entire population. For example, participants need to be grouped based on their countries of origin to model disease spread in the arrival and departure stages of the event. In addition, modeling epidemics in global MGs needs to handle the multiple spatial scales as the participants in these events arrive from different geographical locations and gathered in a specific local space. Even within the setting of the event the participants can go back and forth to different locations either randomly (e.g. audience at Olympic games) or following a specific rituals (e.g. pilgrims at Hajj). Modeling epidemics at MGs requires advanced spatial-temporal simulations that allow the introduction of prevention and mitigation measures. Therefore, it is important to capture and represent the interactions and movements of participants at these different spatial-temporal levels.

5 Summary

Global mass gatherings can pose public health threats on a large scale. Thus, it is important to apply computational epidemic modeling to estimate potential global epidemics. Based on the comprehensive review of the literature presented in this paper, we were able to summarize the key data requirements to model disease spread in these events. While there are several required data related to both the participants attending the event and the event itself, the most important aspect is the incoming and outgoing travel patterns before and after the event. Travel patterns of the participants can be used to predict the path and the dynamics of a potential global epidemic. There is a need for more advanced studies in the context of modeling disease spread in MGs. These studies should provide reliable methods to determine or adequately estimate the incoming and outgoing travel patterns, the event size and setting, and the characteristics of the expected population. These studies should also identify the best methodologies to utilize and integrate these data into disease-spread models. The diversity of the data sources provides opportunities to apply big data methods to assist developing epidemic models in the context of MGs. These challenges and opportunities include large data gathering, simulating enormous heterogeneous populations, approximate the underlying contact patterns among the participants using different data sources, and capturing the multiple spatial-temporal layers.

References

1. World Health Organization: Communicable Disease Alert and Response for Mass Gatherings: Key Considerations. WHO, Geneva (2008)
2. Memish, Z.A., Stephens, G.M., Steffen, R., Ahmed, Q.A.: Emergence of medicine for mass gatherings: lessons from the Hajj. Lancet. Infect. Dis **12**(1), 56–65 (2012)
3. Ebrahim, S.H., Memish, Z.A., Uyeki, T.M., Khoja, T.A., Marano, N., McNabb, S.J.: Pandemic H1N1 and the 2009 Hajj. Science **326**(5955), 938 (2009)
4. Al-Tawfiq, J.A., Smallwood, C.A., Arbuthnott, K.G., Malik, M.S., Barbeschi, M., Memish, Z.A.: Emerging respiratory and novel coronavirus 2012 infections and mass gatherings/infections respiratoires émergentes, nouveau coronavirus 2012 et rassemblements de masse. East. Mediterr. Health J. **19**, S48 (2013)
5. Tabatabaei, S.M., Metanat, M.: Mass gatherings and infectious diseases epidemiology and surveillance. Int. J. Infection **2**(2) (2015)
6. Gautret, P., Parola, P., Brouqui, P.: Risk factors for H1N1 influenza complications in 2009 Hajj pilgrims. Lancet **375**(9710), 199–200 (2010)
7. Al-Jasser, F.S., Kabbash, I.A., AlMazroa, M.A., Memish, Z.A.: Patterns of diseases and preventive measures among domestic hajjis from Central Saudi Arabia. Saudi Med. J. **33**(8), 879–886 (2012)
8. Ziyaeyan, M., Alborzi, A., Jamalidoust, M., Moeini, M., Pouladfar, G.R, Pourabbas, B., et al.: Pandemic 2009 influenza A (H1N1) infection among 2009 Hajj pilgrims from Southern Iran: a real-time RT-PCR based study. Influenza Other Respir.Viruses **6**(6), e80–e84 (2012)
9. Kandeel, A., Deming, M., Elkreem, E.A., El-Refay, S., Afifi, S., Abukela, M., Earhart, K., El-Sayed, N., El-Gabay, H.: Pandemic (H1N1) 2009 and hajj pilgrims who received predeparture vaccination, Egypt. Emerg Infect Dis. **17**(7), 1266–1268 (2011)
10. Gautret, P., Vu Hai, V., Sani, S., Doutchi, M., Parola, P., Brouqui, P.: Protective measures against acute respiratory symptoms in French pilgrims participating in the Hajj of 2009. Journal of travel medicine **18**(1), 53–55 (2011)
11. Gutiérrez, I., Litzroth, A., Hammadi, S., Van Oyen, H., Gerard, C., Robesyn, E., Bots, J., Faidherbe, M.T., Wuillaume, F.: Community transmission of influenza A (H1N1) V virus at a rock festival in Belgium, 2–5 July 2009. Euro surveillance: bulletin Europeen sur les maladies transmissibles = Eur. Commun. Dis. Bull. **14**(31), 2202–2206 (2009)
12. Abubakar, I., Gautret, P., Brunette, G.W., Blumberg, L., Johnson, D., Poumerol, G., Memish, Z.A., Barbeschi, M., Khan, A.S.: Global perspectives for prevention of infectious diseases associated with mass gatherings. Lancet. Infect. Dis **12**(1), 66–74 (2012)
13. Gundlapalli, A.V., Rubin, M.A., Samore, M.H., Lopansri, B., Lahey, T., McGuire, H.L., Winthrop, K.L., Dunn, J.J., Willick, S.E., Vosters, R.L., Waeckerle, J.E.: Influenza, winter olympiad, 2002. Emerg. Infect. Dis. **12**(1), 144 (2006)
14. Jorm, L.R., Thackway, S.V., Churches, T.R., Hills, M.W.: Watching the games: public health surveillance for the Sydney 2000 olympic games. J. Epidemiol. Community Health **57**(2), 102–108 (2003)
15. Steffens, I., Martin, R., Lopalco, P.: Spotlight on measles 2010: measles elimination in Europe— a new commitment to meet the goal by 2015. Euro. Surveill. **15**(50), 19749 (2010)
16. Steffen, R., Bouchama, A., Johansson, A., Dvorak, J., Isla, N., Smallwood, C., Memish, Z.A.: Non-communicable health risks during mass gatherings. Lancet Infect. Dis. **12**(2), 142–149 (2012)
17. Johansson, A., Batty, M., Hayashi, K., Al Bar, O., Marcozzi, D., Memish, Z.A.: Crowd and environmental management during mass gatherings. Lancet Infect. Dis. **12**(2), 150–156 (2012)
18. Khan, K., McNabb, S.J., Memish, Z.A., Eckhardt, R., Hu, W., Kossowsky, D., Sears, J., Arino, J., Johansson, A., Barbeschi, M., McCloskey, B.: Infectious disease surveillance and modelling across geographic frontiers and scientific specialties. Lancet. Infect. Dis. **12**(3), 222–230 (2012)

19. Tam, J.S., Barbeschi, M., Shapovalova, N., Briand, S., Memish, Z.A., Kieny, M.P.: Research agenda for mass gatherings: a call to action. Lancet. Infect. Dis. **12**(3), 231–239 (2012)
20. Nsoesie, E.O., Kluberg, S.A., Mekaru, S.R., Majumder, M.S., Khan, K., Hay, S.I., Brownstein, J.S.: New digital technologies for the surveillance of infectious diseases at mass gathering events. Clin. Microbiol. Infect. **21**(2), 134–140 (2015)
21. Daley, D.J., Gani, J., Gani, J.M.: Epidemic Modelling: An Introduction. Cambridge University Press, Cambridge (2001)
22. Chowell, G., Nishiura, H., Viboud, C.: Modeling rapidly disseminating infectious disease during mass gatherings. BMC Med. **10**(1), 159 (2012)
23. Khan, K., Freifeld, C.C., Wang, J, Mekaru, S.R., Kossowsky, D., Sonricker, A.L., Hu, W., Sears, J., Chan, A., Brownstein, J.S.: Preparing for infectious disease threats at mass gatherings: the case of the Vancouver 2010 olympic winter games. Can. Med. Assoc. J. **182**(6), 579–583 (2010)
24. Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Colizza, V., Isella, L., Régis, C., Pinton, J.F., Khanafer, N., Van den Broeck, W., Vanhems, P.: Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. BMC Med. **9**(1), 87 (2011)
25. Shi, P., Keskinocak, P., Swann, J.L., Lee, B.Y.: The impact of mass gatherings and holiday traveling on the course of an influenza pandemic: a computational model. BMC Public Health **10**(1), 1 (2010)
26. Cowling, B.J., Chan, K.H., Fang, V.J., Lau, L.L., So, H.C., Fung, R.O., Ma, E.S., Kwong, A.S., Chan, C.W., Tsui, W.W., Ngai, H.Y.: Comparative epidemiology of pandemic and seasonal influenza a in households. N. Engl. J. Med. **362**(23), 2175–2184 (2010)
27. World Health Organization. Taking sex and gender into account in emerging infectious disease programme: an analytical framework. WHO Regional Office for the Western Pacific, Manila (2011)
28. The European Centre of Disease Prevention and Control (ECDC). Seasonal influenza vaccination in EU/EEA, influenza season 2010–11. http://ecdc.europa.eu
29. The European Centre of Disease Prevention and Control (ECDC). Seasonal influenza vaccination in EU/EEA, influenza season 2011–12. http://ecdc.europa.eu
30. Al-Jasser, F.S., Kabbash, I.A., AlMazroa, M.A., Memish, Z.A.: Patterns of diseases and preventive measures among domestic hajjis from Central Saudi Arabia. Saudi Med. J. **33**(8), 879–886 (2012)
31. Khan, K., Arino, J., Hu, W., Raposo, P., Sears, J., Calderon, F., Heidebrecht, C., Macdonald, M., Liauw, J., Chan, A., Gardam, M.: Spread of a novel influenza A (H1N1) virus via global airline transportation. New England journal of medicine **361**(2), 212–214 (2009)
32. Khan, K., Sears, J., Hu, V.W., Brownstein, J.S., Hay, S., Kossowsky, D., Eckhardt, R., Chim, T., Berry, I., Bogoch, I., Cetron, M.: Potential for the international spread of middle East respiratory syndrome in association with mass gatherings in Saudi Arabia. PLoS Currents **5** (2013)
33. Kao, R.R., Haydon, D.T., Lycett, S.J., Murcia, P.R.: Supersize me: how whole-genome sequencing and big data are transforming epidemiology. Trends Microbiol. **22**(5), 282–291 (2014)