Quantifying spatiotemporal heterogeneity of MERS-CoV transmission in the Middle East region: A combined modelling approach

Chiara Poletto [a],[*], Vittoria Colizza [a],[b], Pierre-Yves Boëlle [a]

[a] *Sorbonne Universités, UPMC Univ Paris 06, INSERM, Institut Pierre Louis d'épidémiologie et de Santé Publique (IPLESP UMRS 1136), F75012, 27 rue Chaligny, Paris 75012, France*
[b] *Institute for Scientific Interchange Foundation, via Alassio 11/c, Torino 10126, Italy*

ABSTRACT

MERS coronavirus cases notified in the Middle East region since the identification of the virus in 2012 have displayed variations in time and across geography. Through a combined modelling approach, we estimate the rates of generation of cases along the zoonotic and human-to-human transmission routes and assess their spatiotemporal heterogeneity. We consider all cases notified to WHO from March 2012 to mid-September 2014. We use a stochastic modelling of the time series of case incidence in the Middle East region to estimate time- and space-dependent zoonotic and human-to-human transmission parameters. The model also accounts for possible lack of identification of secondary transmissions among notified cases. This approach is combined with the analysis of imported cases out of the region to assess the rate of underreporting of cases. Out of a total of 32 possible models, based on different parameterisation and scenario considered, the best-fit model is characterised by a large heterogeneity in time and across space for both zoonotic and human-to-human transmission. The variation in time that occurred during Spring 2014 led to a 17-fold and 3-fold increase in the two transmissions, respectively, bringing the reproductive rate to values above 1 during that period for all regions under study. The model suggests that 75% of MERS-CoV cases are secondary cases (human-to-human transmission), which is substantially higher than the 34% of reported cases with an epidemiological link to another case. Overall, estimated reporting rate is 0.26. Our findings show a higher level of spatial heterogeneity in zoonotic transmission compared to human-to-human, highlighting the strong environmental component of the epidemic. Since sporadic introductions are predicted to be a small proportion of notified cases and are responsible for triggering secondary transmissions, a more comprehensive understanding of zoonotic source and path of transmission could be critical to limit the epidemic spread.

1. Introduction

The Middle East respiratory syndrome coronavirus (MERS-CoV) was detected for the first time in 2012 in the Arabian Peninsula and since then it has been the source of global health concern (WHO). The disease can be severe, it is characterised by sporadic emergence of cases and associated clusters in a restricted geographical area, the Middle East, but few cases also travelled internationally and were diagnosed worldwide (ECDC). More recently it also led to a large local outbreak in South Korea, following a case importation from the Middle East region (WHO and MERS-Cov).

Infection control outbreak investigations and observational studies have highlighted important aspects of the epidemiology of the disease. Both zoonotic introductions and human-to-human transmissions are responsible for new infections. Dromedary camels have been found to be intermediary hosts for the virus and may represent the source for zoonotic infection (Muller et al., 2015; Alagaili et al., 2014). Human-to-human transmission can occur following close and prolonged contacts and it has been identified as the main mechanism responsible for the documented hospital outbreaks of Al-Hasa (Assiri et al., 2013) and Jeddah (Drosten et al., 2015). Modelling studies highlighted the subcritical nature of the epidemic (Chowell et al., 2014; Poletto et al., 2014; Breban et al., 2013; Cauchemez et al., 2014; Majumder et al., 2014) and few of them also characterised the rate of zoonotic transmission (Poletto et al., 2014; Cauchemez et al., 2014). A population-based serological investigation recently provided evidence for substantial

under-detection of cases possibly due to sub-clinical forms of the disease (Muller et al., 2015).

Despite these advances, there are still many gaps in knowledge regarding MERS epidemiology and ecology. Geographical variation in the epidemic has rarely been considered in these studies, even though epidemic data show a clear spatial heterogeneity across countries in the Middle East region and across provinces of Saudi Arabia. Occasional sparks in activity have been observed as well. A particularly large increase in the number of cases was reported during Spring 2014, with a maximum of 133 weekly cases at the end of April (epidemiological week 2014-17), confined mainly in the provinces of Riyadh and Makkah in Saudi Arabia (WHO).

The observed strong spatial and temporal heterogeneities characterising a persistent yet subcritical epidemic pose a set of challenges for the comprehensive understanding of MERS-CoV circulation in the Middle East region. Whether this behaviour results from a combination of different transmission modes or seasonal effects or other mechanisms is still unclear. Here we aim to quantify the spreading potential associated to zoonotic and human-to-human transmission routes and characterise their temporal and geographical variation, to uncover possible variations in the epidemiology of the disease. We propose a combined approach, akin to the one introduced in (Poletto et al., 2014), based on the joint analysis of imported cases out of the Middle East region and incidence of cases within the region. By maximising the use of available information the model is able to estimate the transmission parameters for each route and their spatiotemporal variation, account for unidentified human-to-human transmission among detected case, and assess the rate of case detection.

2. Methods

2.1. Data

Data consisted of all cases notified to WHO with onset between the beginning of the epidemic in March 2012 (specifically week 2012-12) and mid-September 2014 (i.e. week 2014-38), retrieved as of March 2015 from (Rambaut, 2013). Cases arising in the Middle East region (Saudi Arabia, Qatar, Oman, Kuwait, Jordan, United Arab Emirates, and Yemen) and cases imported to Western European countries and North America were analysed separately. The dataset used for the analysis did not include 113 cases occurring between May 2013 and May 2014 identified through retrospective review and reported by WHO on June 26 (MERS-Cov) for which information of neither date of onset or date of hospitalisation was available. Completeness of the dataset used in both temporal and geographical information is addressed in Section 1 of the Supporting Information.

In the Middle East region, cases were grouped according to location at the province level in Saudi Arabia (11 over the 13 provinces experience cases and were considered in the analysis) and at the country level otherwise, finally yielding 17 regions (11 provinces in Saudi Arabia, plus the remaining 6 countries of the Middle East region, i.e. Qatar, Oman, Kuwait, Jordan, United Arab Emirates, and Yemen). For each region we computed the epidemic curve using dates of onset. As this date was missing in approximately half the cases, imputation was used based on hospitalisation or notification dates as explained in the next subsection.

A proportion of 34% of cases reported to WHO were described as "secondary" when they were epidemiologically linked to another case. Accordingly, we refer in the following to cases arising from human-to-human transmission as "secondary cases" and to all others as "primary cases", i.e. of zoonotic origin. Lack of information on the identification of secondary transmission is also considered in the model.

We finally considered the imported cases in Western Europe and North America ($n = 9$), as these regions have set up high sensitivity surveillance systems to detect importations. We included Austria, Belgium, Denmark, Finland, France, Germany, Iceland, Ireland, Italy, Liechtenstein, Luxemburg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, Greece, Canada and the United States as possible destinations out of the Middle East. In this area, 1 case with travel history to Middle East was notified in France, Germany, Greece, Italy and the United Kingdom and 2 cases in the Netherlands and the United States over the period under study.

2.2. Reconstruction of incidence time-series

Incidence of onset time-series in the 17 Middle East regions were reconstructed as follows. If hospitalisation date t_h was available, the onset date was imputed at $t_h - \Delta t_h$ where Δt_h was sampled from the distribution of time from onset to hospitalisation, determined from other cases hospitalized in the same period (i.e. [$t_h - 30$ days, $t_h + 30$ days]); if only the notification date t_r was known, the onset date was imputed at $t_r - \Delta t_r$ where Δt_r was sampled from the distribution of time from onset to notification, determined from other cases notified in the same period, as before. Cumulative distributions from onset to hospitalisation and from onset to notification are reported in the Supplementary Information. This procedure was repeated to yield 20 epidemic curves: there was an average 707 cases (range 704–712 cases depending on imputation) with onset in the period considered for analysis. The overall epidemic curve profile was little affected by imputation (see Supplementary Information). Overall, the five regions reporting the most cases were Makkah (245 cases), Riyadh (223 cases), Eastern Province (68), United Arabia Emirates (65) and Al Medinah (42).

2.3. Model

We used a two-step approach to estimate the reporting rate ρ, the rate of sporadic generation of cases from non-human source $p_{sp}^r(t)$ and the reproduction ratio $R^r(t)$ as a function of time t in each region r. The first step included modelling the incidence time series in the Middle East region, the second step used importation of cases and a stochastic data-driven model for spatial diffusion of the epidemic worldwide. A schematic representation of the analysis is reported in Fig. 1. All parameters and variables are listed and described in Table 1.

Table 1
Variables and corresponding descriptions.

| Variable	Description
$n^r(t)$	Total number of cases in region r at time t
$D^r(t)$	Number of notified cases in region r at time t
ρ	Reporting rate, $D^r(t) = \rho n^r(t)$
N_r	Population of region r
$d_s^r(t)$	Number of secondary transmissions among all notified cases $D^r(t)$ in region r at time t
$s^r(t)$	Number of notified secondary cases in region r at time t
$p_{sp}^r(t)$	Rate of generation of sporadic cases in region r at time t
$q_{sp}^r(t)$	Rate of generation of sporadic cases in region r at time t obtained from notified cases, $q_{sp}^r(t) = \rho p_{sp}^r(t)$
$R^r(t)$	Reproduction ratio for human-to-human transmission in region r at time t
α_r	Geographical variation of the rate of generation of sporadic cases in region r, $p_{sp}^r(t) = \alpha_r p_{sp}(t)$
β_r	Geographical variation of the reproduction ratio for human-to-human transmission in region r, $R^r(t) = \beta_r R(t)$
$q_{sp,1}, q_{sp,2}$	Baseline level and peak level of the spline function assumed to model $q_{sp}^r(t)$
R_1, R_2, R_3	Three levels of the stepwise function assumed to model $R^r(t)$
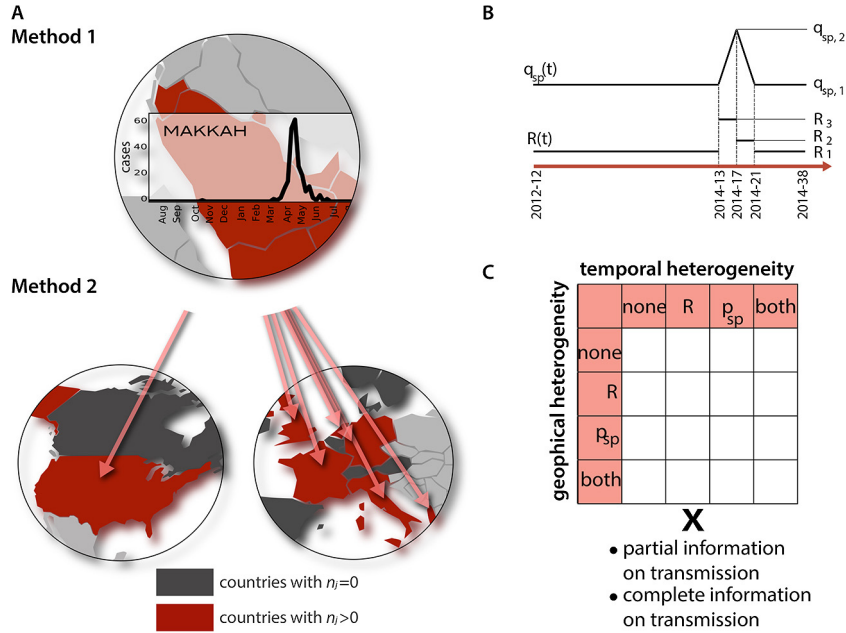
Fig. 1. Model A: Scheme of the 2-step approach. Step 1 is based on the fit of 20 imputed epidemic curves for each of the 17 regions in the Middle East (only one curve is displayed for the sake of visualisation corresponding to the region experiencing the largest number of cases). It allows model selection and estimation of $R^r(t)$ and $q^r_{sp}(t) = \rho p^r_{sp}(t)$. Step 2 is based on the fit of imported cases in Western Europe and North America and allows estimating ρ. **B:** Scheme of the functional forms assumed for $R(t)$ and $q_{sp}(t)$ when temporal heterogeneity (either for one of the parameters or both) is considered in the model. Parameters $q_{sp,1}$, $q_{sp,2}$, R_1, R_2, R_3 are estimated. **C:** Combination of parameters yielding the 32 models for exploration.

First step: Cases were grouped by week, corresponding to the mean generation time (Assiri et al., 2013; Poletto et al., 2014). In each region r, incidence was modelled as the superposition of sporadic cases at a region-specific time-varying weekly rate $p^r_{sp}(t)$ and of cases caused by human-to-human transmission with region-specific time-varying reproduction ratio $R^r(t)$. More precisely, we assumed $p^r_{sp}(t) = \alpha_r p_{sp}(t)$ and $R^r(t) = \beta_r R(t)$, where α_r and β_r are region specific rescaling factors of sporadic generation of cases and reproductive ratio, respectively. We then expressed the expected value of total number of cases $n^r(t)$ in region r at time t as $E(n^r(t)) = \alpha_r N_r p_{sp}(t) + \beta_r R(t-1) n^r(t-1)$ where N_r is the population of the region. As a sensitivity analysis, we also explored a model where the number of cases depended on incidence in the three weeks before (i.e. generation time distribution spanning 3 weeks).

We assumed a constant and uniform reporting rate ρ, so that the number of notified cases $D^r(t)$ in region r at time t was $D^r(t) = \rho n^r(t)$. We further assumed that $D^r(t)$ was Poisson distributed. The observed rate of generation of sporadic cases based on notifications is then defined as $q^r_{sp}(t) = \rho p^r_{sp}(t)$.

As indicated in the Data section, a number $s^r(t)$ of cases among the $D^r(t)$, were identified as secondary. We defined two scenarios for interpreting this information. In the *complete information on transmissions* scenario, $s^r(t)$ was assumed to be equal to the actual number of secondary transmissions among notified cases $d^r_s(t)$, $s^r(t) = d^r_s(t)$; in the *partial information on transmissions* scenario, $s^r(t)$ was taken as a lower bound of this number: we thus assumed that some cases not described as secondary could nevertheless have resulted from unidentified human-to-human transmission, i.e. $d_s(t) \geq s^r(t)$.

In the *complete information on transmissions* scenario, we computed in each region the probability of having $D^r(t)$ notified cases, among which $s^r(t)$ human-to-human transmissions: $P(d^r_{sp}(t) = D^r(t) - s^r(t)\&d_s(t) = s^r(t))$ where $d_{sp}(t)$ are the notified cases arising from sporadic generation. Conditional on $D^r(t)$, assuming the same reporting rate ρ for secondary and primary cases, the distribution

of $d^r_s(t)$ was binomial with probability $\beta_r R(t-1)D^r(t-1)/E(D^r(t))$, so that

$$P(d^r_{sp}(t) = D^r(t) - s^r(t)\&d_s(t) = s^r(t))$$
$$= e^{-E(D^r(t))} \frac{(\beta_r R(t-1)D^r(t-1))^{s^r(t)}}{s^r(t)!} \frac{(N^r \alpha_r q_{sp}(t))^{D^r(t)-s^r(t)}}{(D^r(t) - s^r(t))!}.$$

For the partial information on transmissions scenario, we computed

$$P(d^r_s(t) = D(t) - s^r(t)\&d_{sp}(t) \geq s^r(t))$$
$$= e^{-E(D^r(t))} \frac{(\pi \beta_r R(t-1)D^r(t-1))^{s^r(t)}}{s^r(t)!}$$
$$\times \frac{\left((1-\pi)\beta_r R(t-1)D^r(t-1) + N^r \alpha_r q_{sp}(t)\right)^{D^r(t)-s^r(t)}}{(D^r(t) - s^r(t))!}$$

with π the probability that a case generated by human-to-human transmission was correctly described as secondary. In both cases, the final log-likelihood was obtained as a sum over all weeks and regions.

The rescaling factors α_r and β_r, assumed to model the presence of spatial heterogeneity in the zoonotic generation of cases and secondary transmissions, respectively, are described as random parameters log-normally distributed. The absence of geographical variation was also considered, where these parameters were set to 1.

For temporal changes, we modelled $q_{sp}(t)$ as a linear spline and $R(t)$ as a stepwise function (see Fig. 1B) with nodes/jumps at weeks 2012-12, 2014-13, 2014-17, 2014-21, 2014-38 to capture the largest epidemic wave occurring during Spring 2014. $q_{sp}(t)$ was therefore described by 2 parameters (namely $q_{sp,1}$ and $q_{sp,2}$) and $R(t)$ by 3 parameters (R_1, R_2, R_3). In the absence of temporal variation, these quantities were constant in time, i.e. $q_{sp}(t) = q_{sp,1}$ and $R(t) = R_1$.

Considering systematically all parameterisations (possible presence of temporal and/or geographical variation in $p_{sp}(t)$ and $R(t)$) and scenarios (complete and partial information on transmissions) yielded a total of 32 model formulations (Fig. 1C), and each model was fitted to the 20 imputed epidemic curves. We explored the posterior distribution of all parameters of interest using Gibbs sampling (100,000 iterations, 2 chains). Prior distributions for all parameters are reported in the Supplementary Information. We used Deviance Information Criterion (DIC) to identify the best-fitting model, averaging DIC differences over the 20 epidemic curves, and reported posterior means and credible intervals for parameters.

For sensitivity analysis, we shifted the peak of $p_{sp}(t)$ by one week, allowed for a 10-week-long change in $R(t)$ and $p_{sp}(t)$, for a single step increase in $R(t)$ of varying duration, and for dispersion in the generation time duration (see Supplementary Information for details).

Second step: Similarly to (Poletto et al., 2014), we used a global epidemic and mobility model (GLEAM) (Balcan et al., 2009a,b) to fit the number of imported cases in countries out of the Middle East (Fig. 1b).

GLEAM is based on a spatially structured meta-population approach comprising 3362 subpopulations in 220 countries in the world coupled through mobility connections. The model is informed with high-resolution demographic data for 6 billion individuals and multi-scale mobility data including the full air traffic database from the International Air Transport Association (IATA) (IATA) and short-range ground mobility obtained from national commuting data (Balcan et al., 2009b). We modelled MERS-CoV infection dynamics within each subpopulation as the combination of generation of cases from sporadic infections, assumed to be a Poisson process, and a SEIR compartmentalization (susceptible, exposed, infectious, recovered individuals) with generation time of 7.6 days and latency period of 5.2 days (Assiri et al., 2013; Poletto et al., 2014). Transmissibility is parameterised by setting $q_{sp}^r(t)$ and $R^r(t)$, to the estimated values from the best-fit model selected in step 1.

The detection rate ρ was the free parameter of the model. For each value of ρ, GLEAM allowed the generation of stochastic numerical realizations of the MERS-CoV outbreak simulating the local epidemic in the source region and international dissemination events through mobility processes. We thus generated with a Monte Carlo procedure the probability distribution $P_i(n_i)$ of the number n_i of imported MERS-CoV cases in country i out of the seed region. Being all independent importation events, we can define a likelihood function $\mathscr{L}_{\text{imp}}\left(\rho; \left\{n_j^*\right\}\right) = \prod_j P_j\left(n_j^*\right)$, where n_j^* is the empirically observed number of imported cases per country.

3. Results

3.1. Model selection

Altogether, models in the *partial information on transmissions* scenario provided a much better fit than the ones of the *complete information on transmissions* case, suggesting that more transmissions occurred than actually reported (see Supplementary Information for all DIC values).

Table 2 summarises the DIC differences for the 16 models corresponding to the *partial information on transmissions* scenario. The best model included heterogeneity in time and space for both the sporadic generation of cases and the reproductive ratio. The DIC differences show that all the other models fit substantially worse. As expected, the models with no temporal heterogeneity in either $p_{sp}^r(t)$ or $R^r(t)$ had the largest DIC showing that the hypothesis of

Table 2
DIC differences for the 16 models tested in the scenario partial information on transmissions. Differences were averaged over 20 imputed epidemics.

Temporal variation	Geographical variation			
	None	R	p_{sp}	Both p_{sp} and R
None	329	295	136	121
p_{sp}	308	279	54	45
R	249	219	51	41
Both p_{sp} and R	232	206	14	0

constant transmission parameters failed to properly explain the observed incidence profile. Assuming temporal variation in either transmission mode already improved the fit, with a time varying reproduction ratio providing the largest improvement. For spatial heterogeneity, allowing the zoonotic transmission to change between regions was always favoured. The model providing the closest DIC to the best-fit one included temporal change in the rate of sporadic cases and in $R^r(t)$, with spatial heterogeneity only for the zoonotic component.

This conclusion was robust to changes in the models (see sensitivity analysis in the Supplementary Information).

3.2. Spatial and temporal heterogeneity in the transmission modes

Table 3 shows the estimated parameters of the functions $R(t)$ and $q_{sp}(t)$, along with the reporting rate ρ, estimated for the best-fit model. The bivariate posterior distributions of the parameters $q_{sp,1}, q_{sp,2} R_1, R_2, R_3$ are shown in the Supplementary Information. We predicted ρ equal to 0.26 [0.16–0.52] corresponding to an average underreporting ratio of 3.8. For each region the rate of sporadic generation of cases and the reproductive ratio can be computed by multiplying the temporal rescaling functions by the factors $\rho^{-1}\alpha_r$ and β_r, respectively (estimates for α_r and β_r are provided in the Supplementary Information). The bar chart of Fig. 2 presents an overview of the results by focusing on the minimum and the maximum values of the parameters in time for each region, namely $p_{sp,1}^r = \rho^{-1}\alpha_r q_{sp,1}$ and $p_{sp,2}^r = \rho^{-1}\alpha_r q_{sp,2}$ for the rate of zoonotic transmission, and $R_1^r = \beta_r R_1$ and $R_3^r = \beta_r R_3$ for the reproductive ratio. During the epidemic peak of Spring 2014 the generation of sporadic cases was estimated to increase of a factor 17.4. More in detail, for the period of low epidemic activity we predicted a weekly number of sporadic cases on the whole region (including undetected cases) equal to 6.22 [0.56–43.1] that increases up to a maximum of 108.0 [8.0–734.0] during Spring 2014. The mean reproductive ratio estimated for the period of low activity was below one in all regions, ranging from 0.31 [0.06–0.60] to 0.70 [0.50–0.92]. During the four weeks between 2014-13 and 2014-17, it was predicted to increase by a factor 3.3, raising above one for all regions (Fig. 2). It was found to decrease to values below the epidemic threshold and slightly higher than the base values during the second part of the Spring wave (from 2014-17 to 2014-21).

Zoonotic transmission is characterised by larger geographical variations with respect to human-to-human transmission, as confirmed by the coefficient of variations associated to α_r and β_r

Table 3
Parameter estimates obtained from the best fit.

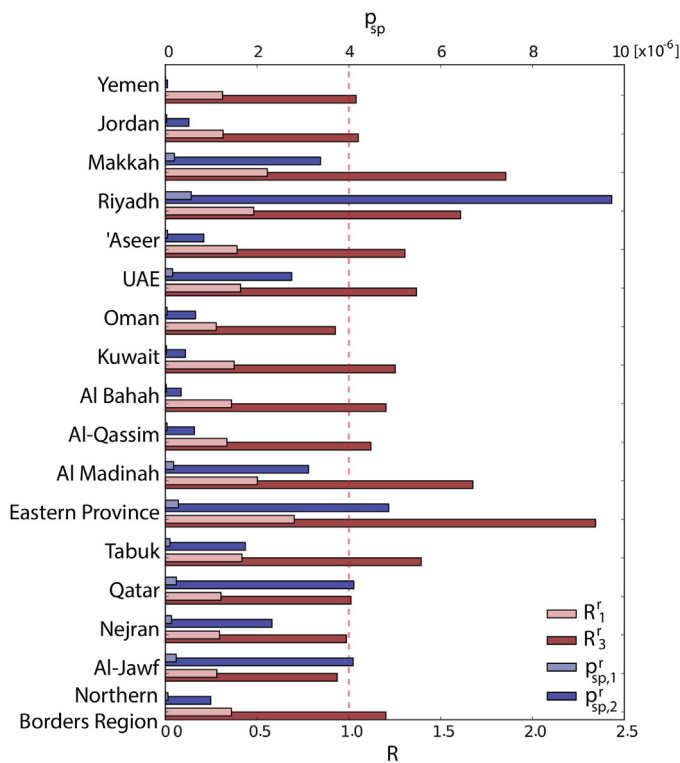|	Parameter estimate
$q_{sp,1}$	$0.026 [0.011–0.056] \times 10^{-6}$
$q_{sp,2}$	$0.45 [0.15–1.07] \times 10^{-6}$
R_1	$0.40 [0.22–0.57]$
R_2	$1.32 [0.72–2.0]$
R_3	$0.60 [0.32–0.85]$
ρ	$0.26 [0.16–0.52]$

Fig. 2. Rate of zoonotic case introduction and reproductive ratio in Middle East regions. Baseline value (minimum, light colour) and Spring 2014 value (maximum, dark colour) are shown in blue for the rate of zoonotic transmission in each region, $p_{sp,1}^r = \rho^{-1}\alpha_r q_{sp,1}$ and $p_{sp,2}^r = \rho^{-1}\alpha_r q_{sp,2}$, and in red for the reproductive ratio $R_1^r = \beta_r R_1$ and $R_3^r = \beta_r R_3$. The dashed red line indicates the threshold value $R = 1$.

equal to 0.99 and 0.20 respectively. The largest reproductive ratios were obtained for the Eastern Region (2.3 [1.4–3.7] during the maximum epidemic activity), Makkah (1.8 [1.3–2.4]), Al Medinah (1.7 [1.0–2.6]) and Ryiadh (1.6 [1.1–2.2]), while the smallest values of the parameter were obtained for Al Jawf (1.0 [0.2–2.1]), Oman (1.0 [0.2, 2.1]). The probability of zoonotic transmission was the highest in the provinces of Riyadh ($p_{sp,2}^r$ equal to 9.5 [3.5–18.3] $\times 10^{-6}$), Eastern Region (4.7 [2.0–8.4] $\times 10^{-6}$) and Qatar (4.1 [1.3–8.5] $\times 10^{-6}$), and the smallest for Al Bahah (0.34 [0.06, 1.15] $\times 10^{-6}$) and Yemen (0.05 [0.01, 0.14] $\times 10^{-6}$).

Results of the sensitivity analysis showed that the posterior distributions of the parameters were little affected by arbitrary and simplifying modelling choices (date of the peak, width of the increase, span of the generation time distribution – see Supplementary Information).

3.3. Assessment of the epidemic situation

Fig. 3 shows the predicted number of cases aggregated over the whole period broken down by type of transmission, obtained for the best-fit model. Results show a larger fraction of secondary cases out of the total number than reported. In WHO reports, secondary cases amount to 34% of all reported cases, with regional levels ranging between 0% and 57%. According to model predictions, this proportion was 75%, with values in the most affected regions equal to 89% in Makkah (vs. the reported 28%), 76% in Eastern Region (vs. 50%), 75% in Al Medinah (vs. 36%), 74% in UAE (vs. 54%) and 69% in Riyadh (vs. 26%). Both the total number of cases and the ratio of secondary to primary cases were geographically heterogeneous.

Fig. 4 shows predicted time series of incidence and proportion of human-to-human transmission cases for the Middle East region as a whole and for the five regions with the most intense

epidemic activity. Periods with the highest levels of incidence were characterised by an increase in the proportion of secondary cases above 50%, indicating an increase in human-to-human transmission above the epidemic threshold. In particular Spring 2014 saw an overall increase in the proportion of human-to-human transmission cases from 52% to 74%. At the regional level the incidence patterns were diverse. Peaks in proportion of secondary cases with values above 95% were reached in Makkah (during the Spring 2014 wave) Al Medinah (in September 2013 and May 2014) and Eastern Province. The province of Riyadh, despite a large incidence, never saw a proportion of transmission above 80%.

4. Discussion

We presented here an assessment of the relative role of human-to-human vs. zoonotic transmission on the dynamics of MERS-CoV in the Middle East peninsula accounting for temporal and geographical variability. We designed a combined modelling approach in order to make optimal use of the available information and provide a comprehensive description of the epidemic. The comparison among a wide set of models highlighted the following characteristics of the MERS-CoV epidemiology.

First, many human-to-human transmission events likely went unidentified, as shown by the better fit of the set of scenarios considering *partial information on transmissions* compared with the ones assuming *complete information on transmissions*. Indeed, human-to-human transmission was reported in 34% of the cases, but the best-fit model predicted that this proportion was up to 75%. Not accounting for the imperfect information on secondary cases led to under-estimating human-to-human transmission. Interestingly, this larger proportion of human-to-human transmission predicted by the best-fit model is consistent with the fact that only few contacts between known sources of zoonotic transmission, such as dromedary camels, and humans have been reported among MERS-CoV patients (Gossner et al., 2014) and often no such contacts were found for infections reported as primary. Our results show that a large part of human-to-human transmission may have went unidentified in the data, as previously suggested (Drosten et al., 2015). The proportion of unidentified human-to-human transmission was the largest when epidemic activity increased, which could be explained by the more difficult assessment of the origin of cases in presence of a large number of cases. Indeed, while in the reported data the proportion of secondary cases did not change substantially during the Spring 2014 period compared to the rest of the time (34% vs. 33%), our model showed that this percentage increased from 52% to 74% in Spring 2014. This sharp increase has previously been described in an observational study and attributed mainly to nosocomial transmissions (Drosten et al., 2015). The predicted value in the low activity period (52%) is similar to the value reported for the period between September 2012 and October 2013 (59%) (The WHO MERS-CoV Research Group, 2013).

Our analysis provided further insight in the Spring 2014 increase. We found that this epidemic wave required a substantial increase in both the rate of introduction of the virus and human-to-human transmission. We obtained indeed a rise of a factor 17 in the rate of sporadic introductions, and a three times increase in the reproductive ratio going from values below the epidemic threshold (in accordance with previous studies (Poletto et al., 2014; Breban et al., 2013; Cauchemez et al., 2014)) to values above such threshold for all the regions. More frequent sporadic introductions could be related to an increase in the virus circulation in the zoonotic source. A recent study (Wernery et al., 2015) on MERS-CoV spread among dromedary camels shows that the virus produces acute epidemics in calves, often born in Spring. Such outbreaks may cause an increase in the number of primary cases and increased opportunities for subsequent transmission, multiplying
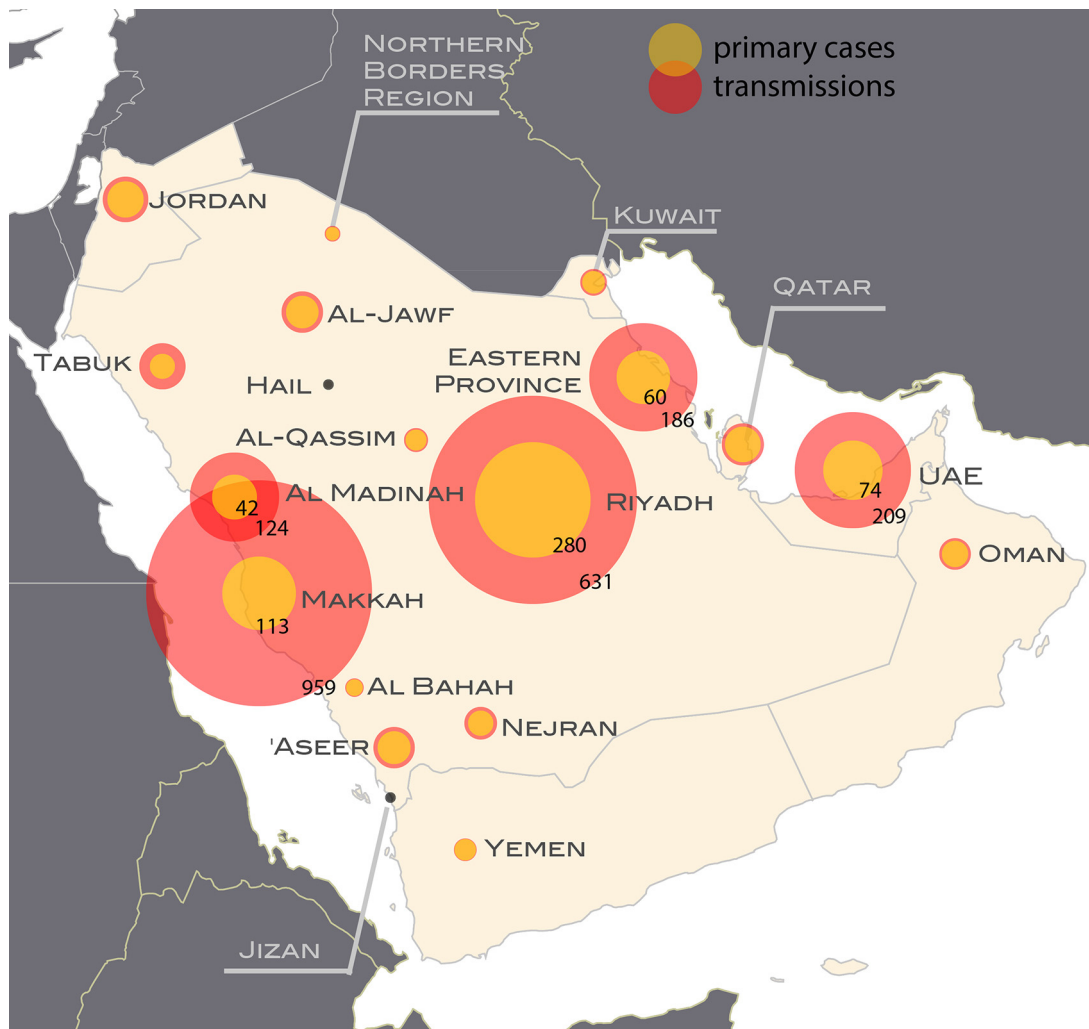
Fig. 3. Cases of zoonotic origin and from human-to-human transmission in each region during the whole period under study. For each region, the areas of the red and yellow portions of the circles are proportional to the number of human-to-human transmissions and zoonotic cases respectively. The actual counts are shown for the five regions experiencing the most cases.

the number of admissions of MERS-CoV cases to hospitals with the possibility of further triggering hospital outbreaks as previously reported (Drosten et al., 2015; Saad et al., 2014; Oboho et al., 2015). Despite a suggestion that the highly seasonal breeding cycle of camels may drive MERS-CoV infections in humans (Wernery et al., 2015), we did not find evidence to explain a seasonal pattern on human-to-human transmission. We tested seasonal patterns for $q_{sp}(t)$ and $R(t)$ to examine if there had been a rise in the number of cases during Spring 2013. Overall, the model fits were not improved (see Supplementary Information). However, this does not rule out that such a seasonal pattern may become more apparent as the number of cases increases. Sensitivity analysis on five alternative models showed that posterior distribution of parameters were little affected by the model details.

Model selection shows that adding geographical heterogeneity in $R(t)$ does not improve substantially the fit, meaning that geographical variation in transmission settings and human-to-human contact behaviour does not play a critical role in the MERS-CoV spreading dynamics within the Arabian Peninsula. On the other hand, zoonotic transmission presented a highest level of heterogeneity. This result points out the strong environmental component of the infection. Such variations indeed should be related to a spatially heterogeneous force of infection induced by the zoonotic source. Recent studies provided information on spatial

density of camels over the Arabian Peninsula (Gossner et al., 2014) and the prevalence of MERS-CoV infection among the animals (Alagaili et al., 2014). Besides these two ingredients, however, the force of infection from camels to human is also determined by human-to-camel contact behaviour. Information on farming practices and their geographical variation would be critical to explain the observed pattern and to inform risk assessment analysis.

The five regions with the highest epidemic activity presented different behaviours. The two largest outbreaks occurred in Makkah and Riyadh. Makkah showed the most intense human-to-human transmission, while the epidemic in Riyadh was characterised by a large number of virus introductions. These findings are consistent with the phylogenetic analysis conducted by (Drosten et al., 2015). Furthermore in accordance with the study in (Majumder et al., 2014) we recovered a higher reproductive ratio in Makkah than in Riyadh (R_3^r equal to 1.9 against 1.6). The epidemic activity in both Al Medinah and UAE shows high transmission and low introductions. Eventually the incidence profile in the Eastern Province is more scattered all over the study period (Assiri et al., 2013).

Another important finding is the confirmation that underreporting in the region is common, as previously reported in other modelling studies (Poletto et al., 2014; Cauchemez et al., 2014), consistently with surveillance works (Saad et al., 2014; Oboho et al., 2015; Memish et al., 2013; Al-Tawfiq and Memish, 2014)
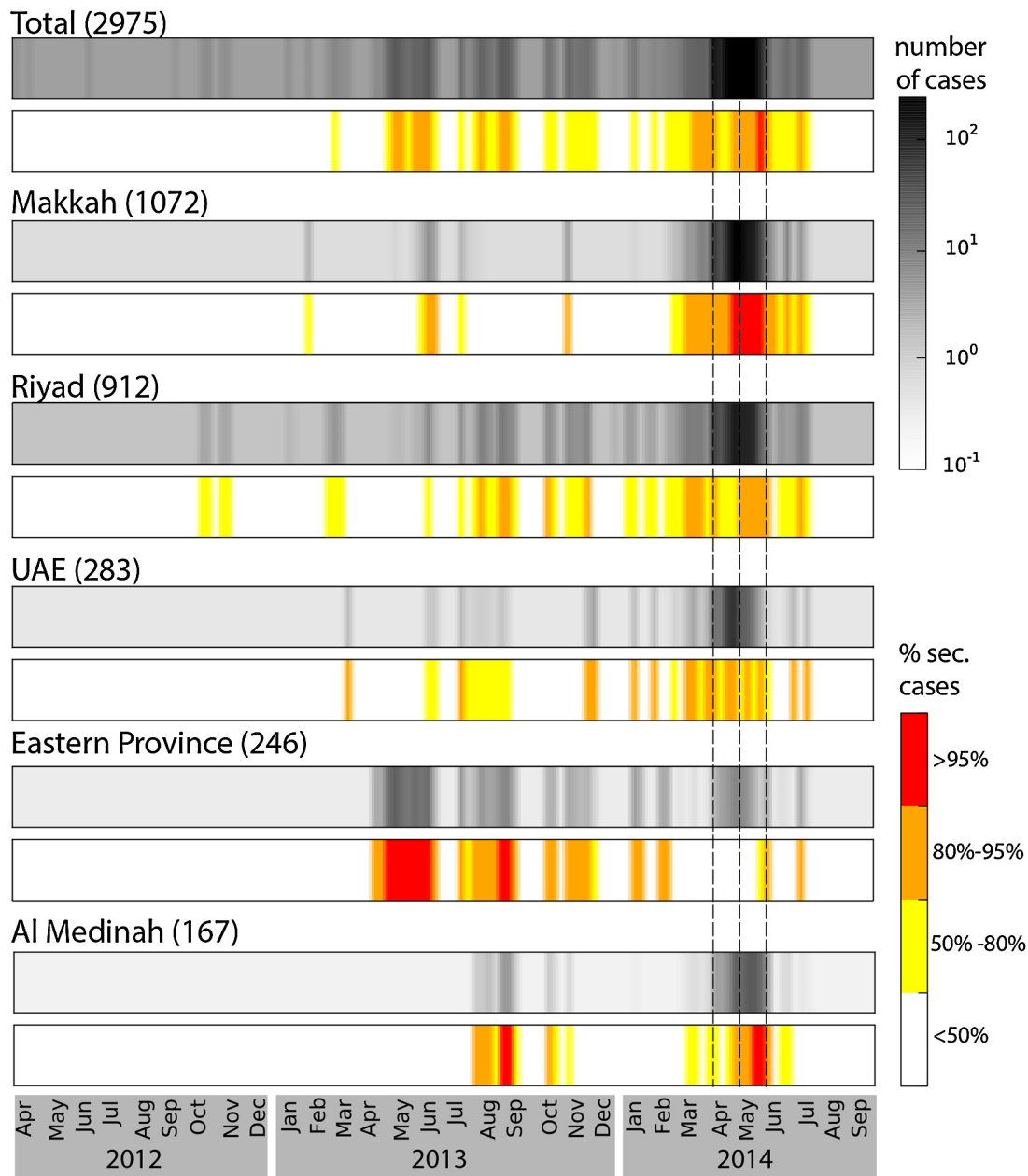
Fig. 4. **Predicted time series of incidence and proportion of secondary cases for the whole Middle East and the five regions with the most cases**. For each region the average number of cases is displayed with the grey-scale colour-map, while values for the proportion of secondary cases are colour-coded (white, yellow orange and red) according to the legend. Average number of cumulative cases predicted by the model is indicated on the top of each colormap. Dashed lines indicate the time periods for increased activity assumed by the model.

and a recent nationwide seroprevalence investigation in Saudi Arabia (Muller et al., 2015). Here, we found that the reported cases amounted to between 16% and 52% of all cases, so that the total epidemic size in the Middle East region as of now could range between ~2500 and ~8000 cases. Cases could go undetected because they are subclinical (Muller et al., 2015; Saad et al., 2014; Oboho et al., 2015; Memish et al., 2013; Al-Tawfiq and Memish, 2014). Assuming mild cases less transmissible then symptomatic ones would have led to a larger estimate for the number of cases, an aspect that was not considered in our model because of lack of information. Current medical knowledge of the disease does not allow a more detailed modelling of the infection natural history and further epidemiological investigation is needed to assess such level of heterogeneity in transmission.

Our modelling approach introduces a seamless transition between analysing cluster of cases and epidemic curves. Most published analyses of the MERS-CoV epidemics focused on the characteristics of case cluster sizes to inform human-to-human transmission (Poletto et al., 2014; Breban et al., 2013; Cauchemez et al., 2014), and would not be adequate to analyse sustained human-to-human transmission. A more recent analysis used trajectory matching based on a deterministic SEIR model (Chowell et al., 2014) that may result in larger inaccuracies for small counts. To overcome these two limitations, we formulated a stochastic birth and death process combining the occurrence of sporadic cases and their offsprings, extending methods for the estimation of reproduction numbers based on secondary cases only to include also sporadic cases (Wallinga and Teunis, 2004). To simplify the

analysis, we considered cases over intervals of the same duration as the mean generation time, but extension to arbitrary generation time distribution is straightforward. We also assumed Poisson variability in incidence, but allowed for over-dispersion by introducing region specific parameters.

Our study is affected by a set of limitations. First, the detection ratio ρ was assumed to be geographically uniform. This may be unrealistic, especially across different countries due to the different surveillance systems. We believe, however, that the assumption of uniform detection rate across the provinces of Saudi Arabia is justified by the national surveillance system and consequently the geographical signature predicted by the model within the region is likely to be genuine. We assumed ρ to be also constant in time. Such assumption was made in other studies referring to the period before Spring 2014 (Chowell et al., 2014); this was justified by the fact that pro-active surveillance was put in place during that period (Memish et al., 2014). The sharp increase in cases during Spring 2014 may have caused an overload in surveillance systems and a decrease in the detection probability. If this is the case, the increase in the zoonotic transmission during the period could be higher than the results provided here. It is important to note, however, that given the model assumptions such bias does not affect the estimated ratio between primary and secondary cases nor the reproductive ratio.

Second, our model assumes transmission probability to be the same across different settings (e.g. households, hospitals). We thus provide unstratified estimates of the reproductive ratio. Current data do not allow a more detailed modelling of the epidemic spread in different settings.

Also, we do not account for the occurrence of super-spreading events. Heterogeneity in transmission was addressed by analysing cluster distribution in Middle East (Kucharski and Althaus, 2015) and local transmission in countries out of the Middle East region following case importation (Nishiura et al., 2015). Results indicate substantial potential for super-spreading that may have contributed to the case insurgence during the Spring 2014 wave.

Eventually, the study restricts to the Middle East region and excludes contiguous countries like Lebanon and Iran, where MERS-CoV cases were detected. This was motivated by the fact that the Arabian Peninsula experienced the great majority of cases and was also the source of importation of cases at the global level (ECDC). Thus the territory represents the focus of major global health concern.

In conclusion, we introduced a new modelling approach applicable to a zoonotic disease in a subcritical/critical spreading regime that is able to disentangle the relative role of the different transmission routes. Applied to MERS-CoV, the model showed that human-to-human transmission is more frequent than expected and high geographical heterogeneity and temporal variation characterise the zoonotic insurgence of cases with bursts that have the potential to trigger outbreaks with intense human-to-human transmission. As such, priority should be given to the control of the zoonotic transmission with a better assessment of the mechanisms at the origin of the observed variation in the generation of cases.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.epidem.2015.12.001.

References

Alagaili, A.N., Briese, T., Mishra, N., Kapoor, V., Sameroff, S.C., de Wit, E., et al., 2014. Middle East respiratory syndrome coronavirus infection in dromedary camels in Saudi Arabia. mBio 5 (2), http://dx.doi.org/10.1128/mBio.00884-14, e00884-14.

Al-Tawfiq, J.A., Memish, Z.A., 2014. Middle East respiratory syndrome coronavirus: transmission and phylogenetic evolution. Trends Microbiol. 22 (10), 573–579, http://dx.doi.org/10.1016/j.tim.2014.08.001.

Assiri, A., McGeer, A., Perl, T.M., Price, C.S., Al Rabeaah, A.A., Cummings, D.A., et al., 2013. Hospital outbreak of Middle East respiratory syndrome coronavirus. N. Engl. J. Med. 369 (5), 407–416, http://dx.doi.org/10.1056/NEJMoa1306742.

Balcan, D., Hu, H., Gonçalves, B., Bajardi, P., Poletto, C., Ramasco, J.J., et al., 2009a. Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility. BMC Med. 7, 45, http://dx.doi.org/10.1186/1741-7015-7-45.

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J.J., Vespignani, A., 2009b. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc. Natl. Acad. Sci. U. S. A. 106 (51), 21484–21489, http://dx.doi.org/10.1073/pnas.0906910106.

Breban, R., Riou, J., Fontanet, A., 2013. Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. Lancet 382 (9893), 694–699, http://dx.doi.org/10.1016/S0140-6736(13)61492-0.

Cauchemez, S., Fraser, C., Van Kerkhove, M.D., Donnelly, C.A., Riley, S., Rambaut, A., et al., 2014. Middle East respiratory syndrome coronavirus: quantification of the extent of the epidemic, surveillance biases, and transmissibility. Lancet Infect. Dis. 14, 50–56, http://dx.doi.org/10.1016/S1473-3099(13)70304-9.

Chowell, G., Blumberg, S., Simonsen, L., Miller, M.A., Viboud, C., 2014. Synthesizing data and models for the spread of MERS-CoV, 2013: Key role of index cases and hospital transmission. Epidemics 9, 40–51, http://dx.doi.org/10.1016/j.epidem.2014.09.011.

Drosten, C., Muth, D., Corman, V., Hussain, R., Al Masri, M., HajOmar, W., et al., 2015. An observational, laboratory-based study of outbreaks of MERS-Coronavirus in Jeddah and Riyadh, Kingdom of Saudi Arabia, 2014. Clin. Infect. Dis. 60 (3), 369–377, http://dx.doi.org/10.1093/cid/ciu812.

European Centre for Disease Prevention and Control (ECDC). Rapid Risk Assessment: Severe respiratory disease associated with Middle East respiratory syndrome coronavirus (MERS-CoV). Stockholm: ECDC; 9 March 2015. Available from: http://ecdc.europa.eu/en/publications/_layouts/forms/Publication_DispForm.aspx?List=4f55ad51-4aed-4d32-b960-af70113dbb90&ID=1274.

Gossner, C., Danielson, N., Gervelmeyer, A., Berthe, F., Faye, B., Kaasik Aaslav, K., 2014. Human–Dromedary Camel interactions and the risk of acquiring zoonotic Middle East respiratory syndrome coronavirus infection. Zoonoses Public Health, http://dx.doi.org/10.1111/zph.12171.

International Air Transport Association (IATA) (accessed on 25 Mar 2014). Available from: www.iata.org.

Kucharski, A.J., Althaus, C.L., 2015. The role of superspreading in Middle East Respiratory Sindrome Coronavirus (MERS-CoV) transmission. Eurosurveillance 20, 25, http://dx.doi.org/10.2807/1560-7917.ES2015.20.25.21167.

Majumder, M.S., Rivers, C., Lofgren, E., Fisman, D., 2014. Estimation of MERS-coronavirus reproductive number and case fatality rate for the Spring 2014 Saudi Arabia outbreak: insights from publicly available data. PLoS Curr. Outbreaks (Dec 18.), http://dx.doi.org/10.1371/currents.outbreaks.98d2f8f3382d84f390736cd5f5fe133c.

Memish, Z.A., Zumla, A.I., Assiri, A., 2013. Middle East respiratory syndrome coronavirus infections in health care workers. N. Engl. J. Med. 369, 884–886, http://dx.doi.org/10.1056/NEJMc1308698.

Memish, Z.A., Al-Tawfiq, J.A., Makhdoom, H.Q., Al-Rabeeah, A.A., Assiri, A., Alhakeem, R.F., et al., 2014. Screening for Middle East Respiratory syndrome coronavirus infection in hospital patients and their health care worker and family contacts: a prospective descriptive study. Clin. Microbiol. Infect. 20, 469–474, http://dx.doi.org/10.1111/1469-0691.12562.

Middle East respiratory syndrome coronavirus (MERS-CoV) – update, 26 June 2014. http://www.who.int/csr/don/2014_06_26/en/.

Müller, M.A., Meyer, B., Corman, V.M., Al-Masri, M., Turkestani, A., Ritz, D., et al., 2015. Presence of Middle East respiratory syndrome coronavirus antibodies in Saudi Arabia: a nationwide, cross-sectional, serological study. Lancet Infect. Dis. 15, 559–564, http://dx.doi.org/10.1016/S1473-3099(15)70090-3.

Nishiura, H., Miyamatsu, Y., Chowell, G., Saitoh, M., 2015. Assessing the risk of observing multiple generations of Middle East respiratory syndrome (MERS) cases given an imported case. Eurosurveillance 20 (27), http://dx.doi.org/10.2807/1560-7917.ES2015.20.27.2118.

Oboho 1, I.K., Tomczyk, S.M., Al-Asmari, A.M., Banjar, A.A., Al-Mugti, H., et al., 2015. 2014 MERS-CoV outbreak in Jeddah—a link to health care facilities. N. Engl. J. Med. 372 (9), 846–854, http://dx.doi.org/10.1056/NEJMoa1408636.

Poletto, C., Pelat, C., Levy-Bruhl, D., Yazdanpanah, Y., Boëlle, P.-Y., Colizza, V., 2014. Assessment of the Middle East respiratory syndrome coronavirus (MERS-CoV) epidemic in the Middle East and risk of international spread using a novel maximum likelihood analysis approach. Eurosurveillance 19 (23), 20824, http://dx.doi.org/10.2807/1560-7917.ES2014.19.23.20824 PMID:24957746.

Rambaut, A., 2013. MERS-cov Spatial Temporal and Epidemiological Information, Available from http://epidemic.bio.ed.ac.uk/coronavirus background (accessed on 31.3.15).

Saad, M., Omrani, A.S., Baig, K., Bahloul, A., Elzein, F., Matin, M.A., 2014. Clinical aspects and outcomes of 70 patients with Middle East respiratory syndrome coronavirus infection: a single-center experience in Saudi Arabia. Int. J. Infect. Dis. 29, 301–306, http://dx.doi.org/10.1016/j.ijid.2014.09.003.

The WHO MERS-CoV Research Group, 2013, November. State of knowledge and data gaps of Middle East respiratory syndrome coronavirus (MERS-CoV) in humans. PLoS Curr. Outbreaks, http://dx.doi.org/10.1371/currents.outbreaks.0bf719e352e7478f8ad85fa30127ddb8, Edition 1.

Wallinga, J., Teunis, P., 2004. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am. J. Epidemiol. 160 (6), 509–516, http://dx.doi.org/10./aje/kwh255.1093.

Wernery, U., Corman, V.M., Wong, E.Y.M., Tsang, A.K.L., Muth, D., Lau, S.K.P., et al., 2015. Acute Middle East respiratory syndrome coronavirus infection in livestock dromedaries, Dubai, 2014. Emerg. Infect. Dis. 21 (6), http://dx.doi.org/10.3201/eid2106.150038.

World Health Organization (WHO), World Health Organization (WHO), Global Alert and Response (GAR), Middle East respiratory syndrome coronavirus (MERS-CoV) Available from: http://www.who.int/csr/disease/coronavirus_infections/en/.

World Health Organization (WHO), Middle East respiratory syndrome coronavirus (MERS-CoV). Summary of MERS statistics in the Republic of Korea (translated from the www.mers.go.kr website) as of 24 June 2015, 9:00. Available from: http://www.wpro.who.int/outbreaks_emergencies/summary.of.MERS.stats/en/ (accessed 24.06.15).