To the editor:

We appreciate the comments from Plachouras et al. on our article published in *Eurosurveillance* a week ago [1,2]. Overall we fully agree with them on both points, i.e., (i) in that there is a need to account for the geographic heterogeneity of the ongoing Ebola epidemic to better understand the transmission dynamics and guide intervention strategies and (ii) in that caution must be exercised to interpret time-dependent changes in the reported coverage of cases captured by the surveillance systems. Here we further highlight these issues by providing feedback from a mathematical modelling point of view.

First, the most recent data points comprising the last three weeks of reported case counts (weeks 35-37) presented by Plachouras et al. were not incorporated in our analysis as these data were not available at the time of preparing our study. Indeed, these additional data points might have changed our interpretation of the most recent trends of the effective reproduction number. Second, our analysis was based on an approximate strategy in line with the available aggregated data. Consequently, we were not able to consider heterogeneous patterns of transmission within each country. With detailed spatial data, we could have detected an apparent slowdown in the incidence influenced by actual decline in incidence at several regions along with a steady increase in Montserrado. With such analysis of spatial data, we would have interpreted the most recent estimate of $R_t$ for Liberia as the result of spatial dilution of differential growth rates by different regions, possible reflection of large local clusters of cases, or the presence of significant reporting delays in the most recent data. Real-time analysis of the ongoing public health crisis in West Africa deserves the consideration of the most detailed, accessible and accurate epidemiological data in order to capture the above-mentioned aspects and explicitly identify regional variations in transmission, which could be key to guide intervention efforts.

We take this opportunity to address two critically important issues in conducting modelling studies using surveillance data subject to limited reporting coverage. First, as discussed in light of our original findings [2], the reported case data are always accompanied by reporting delays. Suppose that the unbiased number of cases and the actual reported number of cases at calendar time $t$ are given by $c_t$ and $r_t$, respectively. Then we have the relationship,

$$c_t = r_t \frac{1}{H_{T-t}}$$

where $H_{T-t}$ is the cumulative distribution function of the reporting delay (of length $T-t$) and $T$ represents the most recent time of observation. This indicates that most recent incidence data might be underestimated (and should be adjusted by $H_{T-t}$). Nevertheless, this might not be a significant issue as long as $H_{T-t}$ is independent of calendar time.

There is a second (and perhaps more serious) issue to consider, i.e., the potential for time-dependent changes in the reporting rate. This is highly relevant to the ongoing Ebola virus disease (EVD) epidemic as the number of new cases has been exponentially growing, which generates pressure on healthcare facilities to assist an extraordinary large number of cases beyond their expected capacity. Let the reporting fraction be $s_t$ at calendar time $t$ which could be estimated by carefully looking into the time-dependent change in the proportion of severe (or fatal) cases among all reported cases [3]. For instance, if the fraction of critically ill cases among total cases increases at a rate $b$ per day, reflecting a decreasing ascertainment rate, we have

$$s_t = \frac{1}{b} s_{t-1}$$

and the unbiased number of cases at $t$, $c_t$, is calculated by dividing the reported number of cases $n_t$ by $s_t$, i.e., $c_t = n_t / s_t$. For instance, a modelling study made a similar
adjustment to analyse data of the influenza A(H1N1)pdm2009 pandemic. In this study, the proportion of hospitalised cases among total reported cases was used as the input data to calculate $s_t$ [3].

It is worth noting that several efforts have already been made to estimate the reproduction number of the ongoing EVD epidemic [2,4,5,6] based on the same publicly available country-wide data of reported cases as in our study.

Potential feedback from modelling studies to surveillance can be summarised as follows: (i) The geographic differences in the evolution of the Ebola epidemic highlighted by Plachouras et al. underscore the need to access high-resolution spatiotemporal data to detect heterogeneous levels in the spatiotemporal dynamics of the epidemic. At the same time, it is critical to exercise caution in the analysis of aggregated time-series data in the presence of significant levels of spatiotemporal heterogeneity. (ii) As a possible indicator of variations in the reporting fraction, monitoring well-defined severe cases would be useful, e.g., hospitalised cases, cases in the state of disseminated intravascular coagulopathy or shock, and deceased cases in order to calculate time-dependent changes in the fraction of the severe cases among the total number of reported cases. It might be also feasible to further account for the time delay from symptoms onset to developing severe manifestations in order to adjust the reporting delay. Surveillance and mathematical modelling are two complementary instruments in the toolbox of epidemiologists. Combining their strengths would be highly beneficial to understand epidemic dynamics and take public health actions. We are keen to contribute further by analysing more detailed epidemiological data of the Ebola epidemic.

Conflict of interest

None declared.

Authors’ contributions

HN and GC drafted and revised the manuscript.

References

4. Althaus CL. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. PLOS Currents Outbreaks. 2014; Sep 2. Edition 1. doi: 10.1371/currents.outbreaks.91af85e0f279e7f729e7056095525b2b