

Classification of community-acquired outbreaks for the global transmission of COVID-19: Machine learning and statistical model analysis

ABSTRACT

Background: As Coronavirus disease 2019 (COVID-19) pandemic led to the unprecedented largescale repeated surges of epidemics worldwide since the end of 2019, data-driven analysis to look into the duration and case load of each episode of outbreak worldwide has been motivated. Methods: Using open data repository with daily infected, recovered and death cases in the period between March 2020 and April 2021, a descriptive analysis was performed. The susceptible-exposed-infected-recovery model was used to estimate the effective productive number (R_t). The duration taken from $R_t > 1$ to $R_t < 1$ and case load were first modelled by using the compound Poisson method. Machine learning analysis using the K-means clustering method was further adopted to classify patterns of community-acquired outbreaks worldwide. Results: The global estimated R_t declined after the first surge of COVID-19 pandemic but there were still two major surges of epidemics occurring in September 2020 and March 2021, respectively, and numerous episodes due to various extents of Nonpharmaceutical Interventions (NPIs). Unsupervised machine learning identified five patterns as "controlled epidemic", "mutant propagated epidemic", "propagated epidemic", "persistent epidemic" and "long persistent epidemic" with the corresponding duration and the logarithm of case load from the lowest (18.6 ± 11.7 ; 3.4 ± 1.8) to the highest (258.2 ± 31.9 ; 11.9 ± 2.4). Countries like Taiwan outside five clusters were classified as no community-acquired outbreak. Conclusion: Data-driven models for the new classification of community-acquired outbreaks are useful for global surveillance of uninterrupted COVID-19 pandemic and provide a timely decision support for the distribution of vaccine and the optimal NPIs from global to local community.