

**Extreme value theory applied to epidemic data:  
on the estimation of return values for major epidemics**

Hans Wackernagel<sup>1</sup>, Geoffroy Brandicourt<sup>1,3</sup>, Huey Chyi Lee<sup>1,2</sup>,  
Fabrice Carrat<sup>3</sup>, Magali Lemaître<sup>3</sup> and Mark Wilson<sup>4</sup>

<sup>1</sup> Geostatistics group, Ecole des Mines de Paris, France

<sup>2</sup> National University of Singapore, Singapore

<sup>3</sup> INSERM Unit 707, Paris, France

<sup>4</sup> Department of Epidemiology, University of Michigan, Ann Arbor, USA

Past influenza pandemics, notably the Spanish influenza (1918-1919) with a death toll of more than 40 million people, and other pandemics such as Asian influenza (1957), Hong Kong influenza (1968), Russian influenza (1977), have greatly impacted humans all around the world. Influenza leads to pneumonia in the more serious cases, and most influenza deaths result from secondary bacterial pneumonia. This occurs more often in the  $> 65$  age group compared to the other age groups. The combined cause-of-death category pneumonia and influenza (P&I) ranks as the seventh leading cause of death in the United States, only to be preceded by heart disease, cancer, stroke, chronic lower respiratory diseases, unintentional injuries and diabetes. Thus, the prediction of future outbreaks of pneumonia and influenza is essential for prevention and control of the magnitude of epidemics.

In this paper we discuss the possibility and associated difficulties in applying methods of statistical analysis of extreme values to US mortality data. We also take care to carefully analyze demographic effects, in particular that of increased aging of the population over the last decades, in order to integrate them into the analysis. Finally, we examine the problem of estimating decadal and centennial return values for large epidemics.