Digitizing and Sharing Public Health Data for Decision Making

Willem G. van Panhuis MD PhD, John Grefenstette PhD, Donald S. Burke MD

Public Health Dynamics Laboratory, University of Pittsburgh

Email: wav10@pitt.edu

Abstract

Although data access and sharing have advanced greatly in various scientific disciplines such as astronomy, genomics, social sciences, earth sciences and ecology, the field of public health is largely lagging behind. Most public health data remains inaccessible for research and decision making due to a wide range of barriers that can be grouped in six categories: technical, incentive related, economic, political, legal, and ethical barriers. In the US, cases and deaths of nationally notifiable infectious diseases have been reported every week by cities and states since 1888 but these data are not available in computable format, making this wealth of data inaccessible for analysis and public health action.

We digitized all weekly US disease surveillance data comprising 87,950,807 individual cases localized in space and time, and analyzed these comprehensive data to quantitatively describe disease control in the USA over the past 120 years. We estimate that national vaccination programs in the US for polio, measles, rubella, mumps, diphtheria, pertussis and hepatitis A prevented over a hundred million cases so far. This data source and the analysis on the impact of vaccination programs will provide an excellent example of the value of sharing historical public health data that were previously inaccessible. We will present this new platform for data sharing in public health and our collaboration with the Taiwan CDC to make public health data maximally available.