

# **Natural Language Processing and Machine Learning for Chronic Disease Management and Prevention: Focus on Asthma**

Wenli Zhang  
wenlizhang@email.arizona.edu  
Department of Management Information Systems  
The University of Arizona

Chronic diseases are the leading causes for decreasing quality of life, hospitalization, disability and death in the United States and all over the world. For a long time, researchers have been seeking ways that promote management and prevention of chronic conditions, to save time, money and energy; to support evidence-based health care decisions and to customize individual patients' disease management plans. Healthcare related Big Data analytical tools have the potential to leverage data from large-scale longitudinal sources for population level chronic disease prevention, as well as to capture trends and propose models for individual level proactive self-management. Nonetheless, the exact role of Big Data analytical tools in the area of chronic disease management has not been fully studied. To take full advantage of Big Data, there is an urgent need to enrich our understanding of Big Data and use it to provide insights for researchers, patients and health providers. By choosing asthma, one of the most serious chronic diseases in the US, as a research case, this dissertation addresses three research questions: (I) How can we use Big Data for asthma surveillance to enable health providers to respond promptly? (II) How can we apply Big Data for asthma risk factors analysis and enhance chronic disease self-management and population level interventions? (III) By identifying smoking as one of the highest population-attributable risk factor, can we use Big Data in evaluating possible substitutes like e-cigarettes? The dissertation comprises of four essays. The first essay seeks to provide an efficient framework to extract signals from social media and make social media data available to answer these questions. The second essay focuses on building a robust Big Data based population level surveillance model that enables health providers to respond to chronic conditions, like asthma, in real time. In the third essay, a framework for comprehensive asthma risk factors analysis is proposed. In the fourth essay I examine behaviors for alleviating chronic disease risks such as smoking. Models, frameworks, and design principles proposed in these essays advance not only health-care research, but also more broadly contribute to design science and predictive modeling research domains.