**Assessing risk factors linked to cardiovascular diseases among US population in the light of socioeconomic and geographic factors**

**Problem:** Cardiovascular diseases (CVD) including are estimated to be the leading cause of death and disability-adjusted life years lost worldwide.1 According to the World Health Organization (WHO) global estimates of mortality and burden of disease, around 17.3 million people died from CVD in 2008 representing 30% of all global deaths.2, 3 Coronary heart disease (CHD) is the second cause of death in people aged under 59 years after HIV/AIDS, reaching the first position in those aged 60 years and above (<http://www.who.int/cardiovascular_diseases>). An estimated one in three U.S. adults (about 71.3 million) has one or more types of CVD. The prevalence increases with advancing age and varies within racial, ethnic, geographic, and sociodemographic groups.

CVD is the most costly disease in U.S. It is estimated that by 2030, more than 23.3 million people will die annually from CVDs.4 By 2030, annual direct medical costs associated with CVDs are projected to rise to more than $818 billion, while lost productivity costs could exceed $275 billion.5 Reducing CVD-associated medical costs would substantially decrease the economic burden on the healthcare system.

The underlying cause of CVDs is atherosclerosis, a process mainly governed by lifestyle factors (risk factors). Identification of such global risk factors has led to the development of risk-prediction algorithms and cardiovascular risk models for men and women. However, they cannot fully explain the excess cardiovascular risk, and at least 25% of all future events occur in individuals with only one of the classical risk factors. Moreover, the prevalence of risk factors is almost as high in individuals without the disease as in patients affected by it. Socio-economic inequalities also have been associated with greater subsequent socio-economic differences in cardiovascular disease. Within high-income countries, low income and low educational status are consistently associated with greater risk of cardiovascular disease.

**Approach:** Previously published findings on the risk factors for CVDs have been reported for either traditional risk or socioeconomic factors, and the role of geographic factors has been poorly studied. Analysis of the whole spectrum of factors in one study has never been reported neither for U.S. nor any other country. U.S. data on CVDs represents a good data material for such a study, since it is available for public and encompasses one-country, high-magnitude variability in the values of socioeconomic and geographic factors. The project will assess modifiable risk factors (obesity, physical inactivity, cholesterol level, hypertension, smoking, diabetes, and nutrition) linked to CVDs (CHD, stroke, and heart attack) among US population (gender, race, and age group) in different states as well as socioeconomic (education, GDP/income, and healthcare spending) and geographic factors (climate and location).

**Proposed Clients:**

*U.S. federal and state health agencies*: Healthcare spending in U.S. is characterized as the most costly compared to all developed countries. The Centers for Medicare and Medicaid reported in 2014 that U.S. healthcare costs were 17.4% GDP in 2013. Therefore, the economic effect of accurate budget planning and reduction and optimization of healthcare spending is hard to overestimate. In addition, increased adherence to clinical and community-level guidelines as well as renewed emphasis on policy, environmental, and lifestyle changes will be crucial for effective prevention and control of CVDs.

*Insurance companies*: Premium rates for policyholders are calculated with the help of risk assessment, which are based on an individual’s key indicators. However, the merits of the numerous composite risk scores currently used for estimating an individual's future risk of CVDs has been debated. Correct risk assessment could save millions of dollars for them.

**Data:**

* Prevalence of traditional risk factors and CVDs among U.S. population (state, gender, race, age group, and geolocation for 2013): <https://catalog.data.gov/dataset/behavioral-risk-factor-data-heart-disease-amp-stroke-prevention>
* Socioeconomic and geographic data: Easily available on the internet

**Deliverables:** The project will include Jupyter notebooks containing code, exploratory data in the form of plots, graphs, or histograms, and a paper. The paper will include a detailed description of the project and the algorithm used as well as the conclusion.

**References:**

1. G Vilahur et all (2014) Perspectives: The burden of cardiovascular risk factors and coronary heart disease in Europe and worldwide. Eur Heart J Suppl (2014) 16
2. R Lozano et all (2010) Lancet, 2012, vol. 380(pg. 2095-2128)
3. Global Status Report on Noncommunicable Diseases, 2010 Geneva World Health Organization
4. CD Mathers, D Loncar (2006) Projections of global mortality and burden of disease from 2002 to 2030, PLoS Med, vol. 3pg. e442
5. <http://www.cdcfoundation.org/pr/2015/heart-disease-and-stroke-cost-america-nearly-1-billion-day-medical-costs-lost-productivity>