Prospects of using large administrative databases in the investigation of complex problems in health

Roberto Fernandes Silva Andrade  
Instituto de Física - Universidade Federal da Bahia

Maria Yury Ichihara, Mauricio L. Barreto  
Centro de Integração de Dados e Conhecimentos para a Saúde - CIDACS-IGM-FIOCRUZ

This work addresses a paradigmatic change in health research motivated by the growing availability of huge amount of data on health and social data of individuals in administrative databases. Indeed, since the construction of the first computers, there has been a steadily increasing capacity of producing, manipulating and storing numerical data. In the last two decades, internet and new kinds of documents (including images, texts, and geo-referenced data) made it possible to share information at a worldwide level, impacting all aspects of daily life, in particular, health. Health research has recently started to use these new resources. One such pioneering experience was launched at the Centro de Integração de Dados e Conhecimentos para a Saúde (CIDACS-IGM-FIOCRUZ). The institutional main goals include receiving data from Brazilian health and social information systems, and in a safety environment, to curate the received databases, to link different databases and to produce large de-identified datasets ($\approx$106 individuals). Based on such datasets, it will be possible to explore and test hypothesis related to determinants of our population health, and provide quantitative support to the use of models to reproduce past events and forecast new policies that can improve the health and social conditions. To tackle the challenges posed by this new paradigm, interdisciplinary knowledge is a fundamental requirement. These include the exploration of new epidemiological designs, data digging methods from data science, statistical analysis of big data, mathematical modelling, and undoubtedly statistical physics and complex systems. We present an overall discussion of the current strategies to develop the CIDACS structure, prospects of scientific problems that can be solved in the new center, as well as preliminary results.