Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

A Model for Early Identification of Chronic Obstructive Pulmonary Disease

Kronborg, Thomas; Hangaard, Stine; Hejlesen, Ole

A Matlab Tool for Organizing and Analyzing NHANES Data

Simon Lebech CICHOSZ[1], Morten Hasselstrøm JENSEN[1,2],
Thomas Kronborg LARSEN[1], Ole HEJLESEN[1]
[1]Department of Health Science and Technology, Aalborg University, Denmark
[2]Steno Diabetes Center North Denmark, Aalborg University Hospital, Denmark

Abstract: Automation of organizing and analyzing NHANES data can provide easier access to data and potentially reducing risk of introducing bias. This study investigates the potential for developing a software for this purpose. MATLAB R2016b was used for transforming and analyzing data from the NHANES. The software was tested successful by analyzing the association between smoking and glucose metabolism in the general population.

1. Introduction

The National Health and Nutrition Examination Survey (NHANES) (1) is a large, stratified, multistage survey conducted by the Centers for Disease Control (CDC) which collects health data on the US population. Data are made publicly available in two-year cycles and the data are categorized into six main areas; dietary, demographics, examination, laboratory, questionnaire, and limited access data.

However, there is an absence of reproducible software and methods that transforms the data from its published online format into an analytic dataset. Even though the NHANES data are publicly available, actually analyzing the data requires a non-trivial amount of background information, data processing and linking. This preprocessing is prone to introducing bias into the data and requires considerable time and resources. We sought to investigate the potential for developing a software for simplifying this process and as an example applying the tool for analyzing the association between smoking and glucose metabolism in the general population.

2. Methods

We developed in MATLAB R2016b a two-level tool. Based on the desired NHANES year-cycles and data variables of interest the first level of the tool will download the relevant data from the public repositories and organize the files locally. The data files are published as a .xpt (SAS XPORT) format, therefore the files are subsequently imported and converted to tables in MATLAB which are then linked based on year-cycle and type of variable. Based on the merged data the second level of the tool is able to analyze data (using an chosen statistical method) and produce tables with patients' characteristics and effect estimates of independent variables in an unadjusted or adjusted analysis. In our study of smoking and glucose metabolism, several years of NHANES (2005-2014) and data types were utilized. From demographics, gender and age data were

collected; from laboratory, fasting glucose and oral glucose tolerance test (OGTT) results were collected; from examination, blood pressure and body mass index (BMI) were collected; and from questionnaire, data on smoking and alcohol habits were collected. The effect of current smoking was assessed in an unadjusted and an adjusted analysis, taking the mentioned covariates into account (gender, age, alcohol usage, BMI).

3. Results

The MATLAB tool was successful in organizing and analyzing data from several year-cycles of NHANES. A total of 12460 participants was identified as currently smoking (n=2575), ex-smokers (n=3032), and never smoking (n=6853). Currently smoking was in the unadjusted analysis associated with higher fasting glucose (effect 3.46 mg/dL, SE:0.72, $p<0.001$). In the adjusted analysis smoking was associated only with lower OGTT (effect -3.94 mg/dL, SE: 1.44, p=0.01)

4. Discussion and conclusion

Automation of organizing and analyzing NHANES data can provide easier access to data and potentially reducing risk of introducing bias. The tool was successfully applied on data from several year-cycles. The test results showed a surprisingly negative association between smoking and OGTT. This association needs to be investigated in more detail.

5. References

[1]   Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2005-2014