Is Information Being Denied to the Scientific Community by the Reductionist Approach to Data Analysis in Stroke Related Clinical Trials?

Cathy Helgason
Department of Neurology
University of Illinois, College of Medicine 912 South Wood Street, Room 855 N
Chicago, IL 60612 (USA)
helgasonkhooby@gmail.com

Massimo Buscema
Semeion Research Center
Of Science of Communications
Via Sersale n. 117
00128 Rome, Italy
m.buscema@semeion.it

Enzo Grossi
Bracco Medical Department
Via 25 Aprile n.4
20097 S.Donato Milanese, Milano, Italy
enko.grossi@bracco.com

Abstract - Background: The African American Anti platelet Stroke Prevention Study was a randomized, double-blind, investigator initiated multi-center trial of 1809 black men and women with recent non cardioembolic stroke. Its goal was to determine the efficacy and safety of two different anti platelet agents, aspirin versus tielopidine, to prevent recurrent stroke, myocardial infarction or vascular death. The results of this study showed no statistically significant difference between the drugs with regards to combined outcome, but a difference approached significance in favor of aspirin for the outcome of stroke. Data regarding the demographics and clinical condition of each patient entered into the trial was collected, in addition to type of stroke. In a different but smaller study, “Influence of Cyclooxygenase-1 and Glycoprotein III a Genotypes on Ex-Vivo Aspirin Response”, the genetic predisposition to aspirin resistance was determined. Again demographic and clinical data were collected on all 59 patients. Statistical analysis suggested that the PTGS1 P17L genotype contributes to aspirin response as measured by ex vivo platelet aggregation studies. Methods: We hypothesized that Auto Contractive Maps, a dynamic system created by Massimo Buscema to create a distance matrix amongst variables of interest would provide information about the relation amongst variables collected in the AAASPS study and Aspirin Response study that not only confirmed but also enriched information provided by standard statistical analysis. The Minimum Spanning tree was extracted from the distance matrix developed by Auto Contractive Maps and compared to Principal Component Analysis. Results: A Minimum Spanning Tree, the most economic way by which to represent the distance between variables, was created for the data set. Connectivity, clustering strength, degree of protection, topological entropy, Delta Hubbness, and Maximally Regular Graph were calculated. Strong links were found between variables in both studies that were missed by Principal Component Analysis. Conclusions: Clinically plausible interactions between variables collected in those patients suffering end point events in the AAASPS study were found using the dynamic non linear mapping method of Auto Contractive Maps. A new interpretation of the importance of genetic predisposition to aspirin response was found in aspirin resistant patients in the smaller clinical study of aspirin response. These connections and new findings were not discovered by PCA. A reductionist approach to data analysis in clinical trials has the potential to deprive the scientific medical community of clinically relevant information.

I. INTRODUCTION

Information from large clinical trials is provided in statistical form based on probability theory. The importance of data mining of large data bases coming from double blind randomized controlled multicenter clinical trials has not been appreciated. Instead, data analysis has taken a reductionist approach limited to probability based statistics. This has especially been the case for studies concerning the diagnosis and treatment of stroke and has meant that the scientific focus has been restricted to the numerical difference in the outcome in two treatment groups when procedures or drugs have been studied. Because the reductionist approach assumes certain relationships amongst variables, it precludes the possibility that hidden associations among variables be discovered. This may have enormous importance when it is recognized that in spite of multiple studies of various agents to treat acute ischemic stroke, none has proven to be statistically significant in beneficial effect for clinical outcome.

Stroke is a complex multi-factorial disease and it is unlikely that a single variable approach will lead to progress in the understanding of the disease. To approach this complex situation, variables of clinical interest should be processed through novel non linear data mining algorithms which are able to extract dynamically hidden information. In this study we applied a novel data mining process to explore the possible association of multiple variables within two different clinical studies: the African American Antiplatelet Stroke Study (AAASPS), a large clinical trial comparing the preventive effect of two different anti platelet agents for recurrent stroke, myocardial infarction and death, and a smaller study, the Aspirin Response Study (ARS), wherein the genetic predisposition to aspirin response as measured by inhibition of