Department of public health sciences
Karolinska Institutet, Stockholm, Sweden

EPIDEMIOLOGY AND STATISTICAL MODELING
In Burn Injuries

AKADEMISK AVHANDLING
som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i Aulan, Plan 2 Norrbackahuset,Karolinska Institutet

Fredagen kl 13.00 den 14 Jan. 2011

av

Homayoun Sadeghi Bazargani

Stockholm 2011

Handledare:
Dr. Reza Mohammadi
Karolinska Institutet
Stockholm

Professor Shrikant Bangdiwala
North Carolina universitet
Karolinska Institutet

Professor Leif Svanström
Karolinska Institutet

Docent Robert Ekman
Karolinska Institutet

Fakultetsopponent:
Professor Peter Byass
Epidemiologi och global hälsa
Umeå universitet, Umeå

Betygsnämnd:
Docent Tahereh Moradi
Karolinska Institutet

Professor Sven-Erik Johansson
Karolinska Institutet

Professor Staffan Janson
Karlsatad universitet

Abstract

An important issue in assessing the epidemiology of injuries, including burns, is the investigation of appropriate methodologies and statistical modeling techniques to study injuries in an efficient and trustworthy manner. The overall aim of this thesis is to analyze epidemiological patterns and assess the appropriateness of supervised statistical models to investigate burn risks and patterns.

This thesis contains four papers: the first two concern descriptive epidemiology of burns in Ardabil Province in Iran, followed by the two methodology papers discussing the applicability and validity of supervised statistical models. Study 1 enrolled 1, 700 minor and moderate burn injury cases, the majority of whom were females and children. Study 2 enrolled 237 burn victims with a slightly higher percentage of males and older patients. The minimum estimated incidence rates were 340 and 13.2 by 100,000 person-years respectively in the first and second studies. Median total body surface area burned was about zero in the first study compared to 15% in the second study. Both studies found the home to be the main injury place , but differed mainly in injury mechanism, agents causing the burn, and the related appliances. Additionally, Study 2 highlighted the two most important injury patterns among women: getting burned while using a camping gas stove or while refilling the fuel chamber of kerosene-burning appliances without first extinguishing them.

In the third study we successfully applied orthogonal projections to latent discriminant analysis(OPLS-DA) to model large numbers of variables in a case-control study and compared it with discriminant analysis done by partial least squares regression (PLS-DA). Prior to fitting the models, the dataset was split into two parts: a training set and a prediction set. Models fitted on the training dataset were later tested for validity in the prediction dataset. The OPLS-DA was compared with PLS-DA for model fitness, diagnostics and model interpretability. Both models suited the data but OPLS-DA was preferable. In Study 4, data from Study 2 were used to investigate the applicability of supervised statistical models in assessing the burn injury patterns, outcomes and their inter-relationship. An unsupervised classification was initially done using principal component analysis. Two separate clusters were observed. Observations were later split into two classes to investigate possible predictors of belonging to each class by PLS-DA as our first supervised statistical model in this study. Based on the results of PLS-DA, the classes were later designated as high-risk burn victims and low-risk burn victims. To assess predictors of TBSA, first a PLS model was fitted. Due to the existence of orthogonal variations, OPLS was also used after PLS regression. Some possible predictors were found to be associated after modeling the natural logarithm of TBSA in the OPLS model. The fitted model could explain 76% of variation in Y. It excluded up to 9 % of orthogonal variation captured in two orthogonal components.

Conclusion: The first basic epidemiological findings of this thesis helped in defining two age-gender based target groups as well as two burn injury patterns of importance for prevention. Based on the results of these studies and considering the potential capabilities of supervised statistical models, the methods discussed in this thesis can potentially be useful to distinguish and define risk indicators, to develop risk prediction scales, assess patterns for understanding the injury mechanisms, detect demography-based target groups and define class-based prevention packages.

Keywords: Burn injuries, Epidemiology, statistical modeling, supervised models, Partial least squares regression, Orthogonal projections to latent structures, Multicollinearity, quantitative content analysis

ISBN 978-91-7457-204-9