From
Department of public health sciences
Karolinska Institutet, Stockholm, Sweden

EPIDEMIOLOGY AND STATISTICAL MODELING
In Burn Injuries

Homayoun Sadeghi Bazargani

Karolinska Institutet

Stockholm 2010

ISBN 978-91-7457-204-9

Dedicated to:

MY WIFE, LOUIZ

Abbreviations & Terminology

Abbreviations

DModX: Distance to model in X space
DModY: Distance to model in Y space
NIPALS: Nonlinear Iterative Partial Least Squares
OSC: Orthogonal signal correction
PLS: Partial least squares [regression]/ projections to latent structures
PLS-DA: Partial least squares discriminant analysis
OPLS: Orthogonal projections to latent structures
OPLS-DA: Orthogonal projections to latent structures discriminant analysis
POSC: Projected orthogonal signal correction
PCA: Principal component analysis
RSD: Residual standard deviation
SIMCA: Soft independent modeling of class analogy
TBSA: Total body surface area[burned]
VIF: Variable inflation factor
VIP: Variable importance in the projection

Terminology

Supervised classification: Also known as discrimination, entails developing algorithms through which observations are assigned to a priori defined categories(like subjects assigned as cases or controls).
Supervised models: In supervised models like PLS or OPLS, other than looking into the internal relationships in the matrices, as in unsupervised models, a main objective is to predict outcome related variables from possible predictors by linking the X and Y matrices
Omics: The English-language neologism omics, informally refers to a field of study in biology ending in -omics, such as genomics or proteomics.
Class-based prevention packages: This is a practical terminology we have used when referring to the prevention packages specifically prepared for population groups, who fall into different classes, based on the observed similarities in a wide range of injury epidemiological features occurring among them.

List of Publications

1- Sadeghi-Bazargani H, Mohammadi R, Svanstrom L, Ekman R, Arshi S, Hekmat S, et al. **Epidemiology of minor and moderate burns in rural Ardabil, Iran.** Burns 2010 Sep;36(6):933-7

2- Sadeghi-Bazargani H, Arshi S, Ekman R, Mohammadi R. **Prevention oriented epidemiology of burn injuries in Ardabil provincial burn center, Iran**. DOI:10.1016/j.burns.2010.09.013. Burns. E-Published

3- Homayoun Sadeghi-Bazargani, Shrikant Bangdiwala, Kazem Mohammad, Reza Mohammadi. **Compared application of the new OPLS-DA statistical model versus partial least squares regression to manage large numbers of variables in a case-control study.** Submitted

4- Homayoun Sadeghi-Bazargani, Shrikant Bangdiwala, Leif Svanstrom, Reza Mohammadi. **Using supervised statistical models to assess injury patterns, outcomes and their interrelationship.** Submitted

…………………………………………………………………………………..

Other publications of the student related but not included in the thesis are given in Appendix1.

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

Contents

1 INTRODUCTION

Injuries are a growing public health problem, substantially affecting nearly every population and every geographical zone in the world. Burns have always been considered as one of the most destructive injuries, causing not only mortalities but also major economic and psychological impacts and long-term somatic sequels as well(3-6).

The approach to burn prevention, to be effective in a particular area, should be based on sound knowledge of etiological patterns of burn injuries and must take into account the geographical variations and socioeconomic differences in burn epidemiology(7;8).

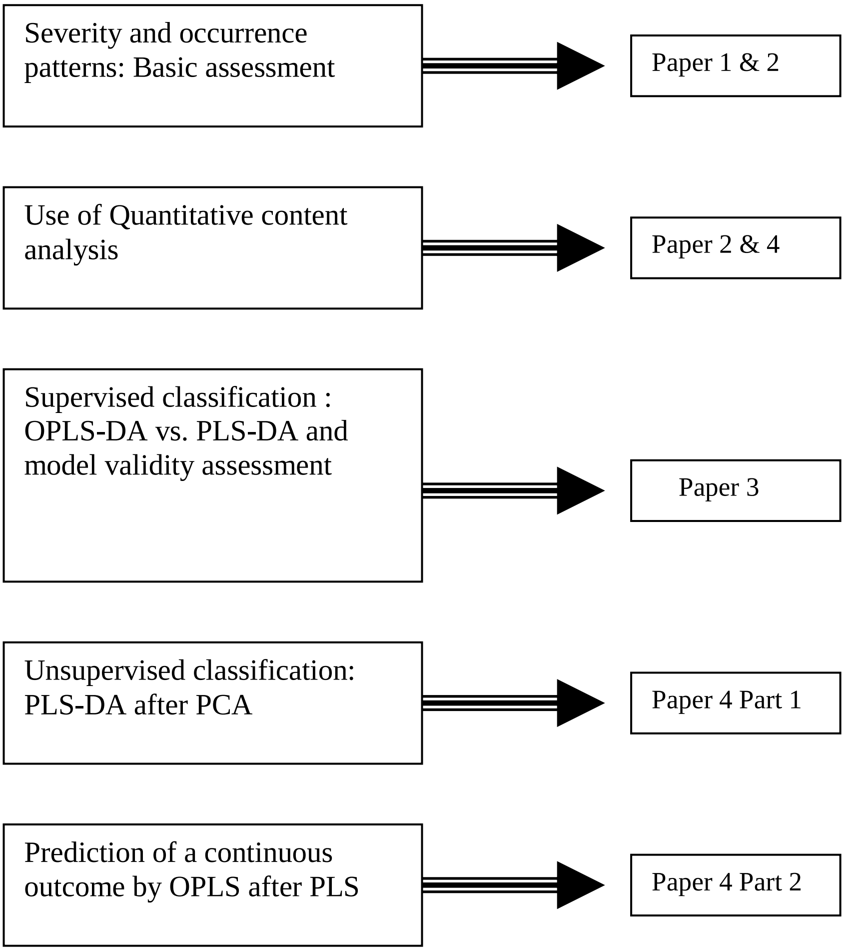
Some preventive measures have been shown to be quite effective in reducing burn injuries. Nevertheless, most of the evidence comes from high-income countries. This is while the patterns and risks of burns can be quite different in low and middle income countries(LMICs), and few of these interventions are readily transferable to LMICs(9).

Considering the variability in etiological and predisposing factors in low and middle-income countries (LMICs), as well as other shortcomings like unavailability of quality databases and registries in these countries, seeking for more compatible research methodologies including statistical methods is a vital issue to be addressed by injuries epidemiologists and statisticians. However, this has been rarely addressed in literature by burn injury researchers. Similarly in Iranian research literature, as will be clarified later in this thesis, most of the studies have tried to define descriptive epidemiology of burns leading to hospitalization and mainly have used information available in hospital files. The effort in this thesis has been to investigate those aspects of burn epidemiology that have been less addressed in previous research and focus has been put on evaluating applicability and validity of supervised statistical methods in analysis of risks and patterns .

2 THESIS OUTLINE

Thesis includes two basic epidemiological studies and two studies about statistical modeling. The main interest in this thesis work was related to the search for an appropriate methodology to define and statistically model some epidemiological characteristics of burns in an Iranian population. In the first study, a suitable screening frame is sought to define the epidemiology of minor and moderate burns in rural areas of Ardabil Province, Iran. The second study investigates similar research questions as Study 1 but for burns leading to hospitalization. However, it also uses a quantitative content analysis of an open-ended question, which describes the pattern of injury occurrence. In this process, new variables were created and analyzed along with information coming from predefined multiple choice questions. The results in Paper 1 and Paper 2 are provided descriptively or just using simple statistical methods, while Study 3 and Study 4 are focused on modeling methodology. Paper 3 is a statistical paper providing only the results of model applicability and diagnostics. It tries to assess the applicability and validity of a new statistical model called orthogonal projections to latent structures discriminant analysis (OPLS-DA) to analyze the case-control study data. The OPLS model is also compared in this study with discriminant analysis done by partial least squares regression. After receiving some preliminary evidence in Study 3, these methods are applied to assess burn injury patterns, outcomes and their interrelationship in Study 4. In addition, it has been of interest in this study to see whether burn victims can be classified in terms of epidemiological characteristics related to injury patterns and outcomes.

Methodological outline of papers



3 BACKGROUND

3.1 BURN EPIDEMIOLOGY

3.1.1 Burns worldwide

Burns are among the leading causes of death and disability worldwide and 265,000 fire-related burn deaths are reported globally each year (10;11). Although many aspects of health status have improved a lot in recent decades, fire-related burn injuries are still responsible for 300,000 deaths and 10 million disability-adjusted life years lost annually worldwide. In the USA burn injuries result in approximately 1,000,000 emergency department visits, 50,000 hospital admissions, and a 5% mortality rate annually and burns are experienced by about 250,000 people in the UK each year (6;12). According to the World Health Organization, 238,000 individuals died of fire-related burns in 2000, and 95% of these deaths occurred in low and middle-income countries (13).

Also non-fatal burns contribute to greater number of cases than fatal burns. However, non-fatal burns epidemiology has not been well elaborated until now.

The discrepancy in incidence rates is obvious among different global zones. Studies recognize burns as being of special importance in India and some African countries and then among Middle East countries. Ultimately, more than 90% of fatal thermal injuries occur in developing countries and in countries with low to moderate incomes (11;14).

3.1.2Burns in Iran

To capture a complete picture of burn epidemiology in Iran as defined in previous research, I started to retrieve the information in a systematic manner as follows.

The information presented here comes from two major sources: The first and the largest source of evidence in this study is from a wide literature review made on articles published regarding burn injuries in Iran.

Original articles published on the epidemiology of burns during the period 2000 to November 2010 were retrieved from four sources as follows

1- Pubmed database
2- Embase database
3- Scopus database
4- SID database, which is the scientific information database developed to index papers published in Iranian journals.

5- A few quality Iranian journals lacking full coverage for the period of publication in the SID database were also checked by hand.

The collected information was finally imported into a Microsoft access software package, and checked for duplications to be removed.

Considering the possible transition of the epidemiological features of burn injuries, papers published earlier than the year 2000 were not selected. Selective key words were used to retrieve the related articles. I also reviewed the references provided at the end of each retrieved article to decrease the chance of missing any papers with unusual keywords and titles. Review articles were not used except in the introduction and discussion parts. The articles about the epidemiology, prevention, mortality, social and psychological consequences of burns were included. Studies regarding treatment, traditional remedies and animal studies were excluded. Duplication of information in two languages was also considered as exclusion criteria. Up to 50 articles were finally chosen to be reviewed.

As the second source, available data from the Iranian National Home Injury Registry Database(INHIRD) during two Iranian calendar years in the period 2001-2003 were used. This database included data for 400,000 injuries registered.

Magnitude of the problem:

In the first national Iranian study on burden of diseases and injuries in 2003, burns came in 13th position among all diseases and injuries (15).According to the preliminary analysis of data belonging to the INHIRD, burns accounted for over 163,000 (41%) of about 400,000 injury cases.

Based on the information from the Tehran Forensic Medicine Council, which represents mortality data automatically collected from 100 hospitals throughout the Iranian capital, burns accounted for 18% of all mortal injury cases. They rank second after traffic injuries, and the proportion of burn injuries is three times the figure for falls or intoxications(16).A mortality study conducted in 10 provinces of Iran during 2000-2001 showed that 14.9 % of all deaths were due to injuries, 80% of them being unintentional injuries, and nearly 27 % of years of lost life were from injuries. This study found that 1% of all deaths were due to unintentional burn injuries(17).Self-immolation incidence rates are also reported to vary from zero to nine per 100,000 person-year over different provinces of Iran(18).

A national community based study of nonfatal injuries [stated to be] conducted in 28 provinces of Iran in 2002-3, found the nonfatal burns[requiring medical care]to have an incidence rate of 10.9 per 100,000 person-years(19). However, it must be added that due to the lack of sufficient community-based information available for minor burns, reliable estimation of the true incidence of all burns in Iran seems to be difficult. Gross estimation of the incidence of minor and moderate burns from a recent study in rural areas of Ardabil Province showed the figure (340 by 100,000 person-years) to be more than 25 times the incidence of burn hospitalization in the Ardabil provincial burn center.

Age distribution in Burns

The average age of burns patients varies from 19 to 35 in different studies (20-32). In most studies this age has been reported as between 21 to 23 years old and the highest average age was found in a study on chemical burns to be 35.3 years (24). Except for two studies, most other studies had found about half the affected patients to be children or under 20 years old (16;24). Totally, the average age of ultimately fatal injuries has been higher and generally is reported to be above 27 years old.

The average age of patients burned through special burning types or agents, has been stated to differ from those affected by conventional types or sources. A mean of 27-31 years has been reported for electrical and lightening injuries (25;33;34) and about 24, for burns during pregnancy(23).

Suicidal burns occur at a higher average age than in accidental cases and this age varies from 24 to 27 years old. They are generally in the age range of 13 to 30. However, such a pattern of burning is rare under 10 years old (35-41).

According to the analysis obtained from INHIRD, average age for all domestic injuries, was 22 years during 2001and 22.5 during 2002. Specifically, for burns, averages were 19.8 and 19.6 during 2001 and 2002 respectively. The averages age was constantly higher for females than for males as respectively recorded 15.5 and 16.1 years for males, vs. 22.6 for females during same years (2001 and 2002).

Agents causing burns

There is a lack of consistency in the description of burning agents throughout different studies. Such contradictions are mostly due to differences in the methodologies or in the populations under study. For example, some evaluations are focused on burn- induced mortalities. Some studies are based on hospital admitted cases and others are community based. Even the results of hospital-based research can emerge differently, depending on the level of sophistication of the services provided by each center. Cultural, geographical, economical and life style differences should also be taken into account. In addition, the roots of a few contradictions can be sought in the diversely used terminology and definitions in reporting about burn injury etiology. For example, in some studies, fuel related burns are separately classified and not included in flame-related burns (42).

Burns related to a gas pipe or cylinder leak or explosion, account for 10% of admitted patients. Flame-burns and scalds are reported to be the most common type of burn injuries, except in those studies conducted on specific types of burns. According to most of the hospital-based and mortality-based studies, flame burns are the most common type of burn injuries (16;21;27;29;30;42-47). This is while community-based studies and INHIRD results, as well as outpatient and minor burn studies, have found scalds to be more common than flame burns (20;31;48;49). One hospital-based study, performed in Shiraz, also revealed hot liquids as the most common cause of burns (26).

However, scalds have been unanimously found to be the predominant form of burns among children (26;45;48;50).Scalds have higher proportions under the age of five. As the age increases, the frequency of scalds decreases and flame burns increase in number. Liquids causing scalds are generally hot water, tea or food (20;49).Though, the steam scalds may account for a substantial frequency of scald injuries, they have not been appropriately described in most Iranian studies.

Results from the Ardabil burn epidemiology study provide more details in this regard (20;49;51). These studies have not only provided details about hot liquid type causing scalds, but also given information regarding liquid container and heating appliances. In contrast to European statistics, the frequency of hot tap water burns in the bathroom is lower in these studies. Electrical burns prevalence, among burn injuries, has been variably reported ranging 2-9 percent (25;30;52;53).

Gender distribution in burns

Most studies have concluded prevalence dominancy among males (16;22;24-27;31;42-45;50;53;54).Few studies reported higher proportions of burns among women(20;49;51). Women dominated in the gender distribution of selfimmolations in all studies conducted in Iran. However, in all of the studies, most of the self-inflicted burn victims were women and female dominance was very high in most studies, often with a ratio as high as three over males (22;30;35-41;46;55-57). One study in Tabriz also reported that 99 % of the self-inflicted burn victims were female (22). The male to female ratio of burned cases obtained from a study performed in Kohkiluye- Boirahmad Province of Iran, were 1.4 and 0.13 for accidental and suicidal burns respectively (30).

A mortality study conducted during 2000-1 showed higher mortality rates for females than males. Regarding the intentional injuries, burns were the only type of injury with female predominance in this study (17).Studies focused on burns with a specific injury mechanism have reported higher male dominance figures in fact as high as 10 times in electrical and chemical burns (24;25;33;34).Burns mechanism and patterns can vary according to gender as well.

Burn mortality

The mortality rate reported by different studies in Iran, varies from 1.4/100,000 to 9.7/100,000 and regardless of study population types, case-fatalities[as interpreted by the author of this thesis] have been variably reported in a range of 2 - 98 % (16;22;24;27;30;32;36;40;41;43;46;47;54;58;59).

The lowest fatality belonged to electrical and chemical burn injuries with a range of 1.7-4.6 percent (24;25;33).The figure regarding electrical injuries seems to be underestimated, because some severe cases of electrical burns result in instant death at the scene and are thus directly referred to forensic medicine. The highest fatality belonged to self-inflicted burns, reaching up to a 75% case-fatality rate in some studies (22;27;30;35;36;60).Higher body surface area burned as reported in fuel related burns may justify the higher fatality rate observed in suicidal burns. These rates fluctuate from 2.1% to 29.2% for non-deliberate burns. TBSA burned shows the strongest association with burn fatality.

Particular conditions, including extremes of age and pregnancy, show dramatically increased burns mortality. An investigation of burns during pregnancy revealed an increased fatality for both maternal and fetal burn which was up to 100% for TBSA above 40% (23).Unlike some other injuries, burn mortalities usually occur not at the scene or during transportation, but rather at the hospital or medical center. However, this proportion is not well defined in the literature.

Time and place for burn injuries

Most of the published studies have not compared burn incidence between rural and urban populations. A few hospital-based studies have just explained the rural/urban ratio of distribution of burns in this regard. However, owing to obvious rural-urban population variation in different provinces of the country, such ratios cannot present an accurate comparison of burn incidence in rural vs. urban areas. A hospital-based study conducted in Hamadan Province of Iran, has found a two-fold higher incidence rate of burn hospitalization for the rural population compared to urban residents (48). Nevertheless, even higher proportions of urban suicidal burns are not reported in Iranian literature, and a recent study in self-immolation also showed a higher incidence in rural areas (16;18;39;44;52). Differences in burn incidence and possibly patterns between the rural and urban population can arise from differences in cultures, environment, life style, cooking habits, beliefs and level of literacy. A community-based mortality study conducted in 10 provinces of Iran during 2000-1, found the mortality rates of burn injuries to be similar both for rural and urban areas equaling four per 100,000 person-years (17).

In most Iranian studies, the home was the most common place at which burns occur, especially for children and women. Suicidal burns also follow the same pattern. The kitchen and living (sitting) room are respectively the most common places where burns occur. Regarding electrical and chemical burns, most of the injuries happened outside of the home.

Burn extent

Burn extent and depth are important factors predicting burn survival. Average TBSA burned has been reported to vary from 1.5% to 65% in different studies. However, this average varied from 30% to 50% among hospital-based studies not focused on any specific group victims and less than five percent for outpatient and minor burns (21-24;26;28;29;31;33;42-44;46;48;49;51;53;60-62).Overall, TBSA percentages were higher among women than men (except for children under 5years old). It was also higher in flame burns than scalds. Average TBSA was generally over 60% in cases resulted in death, and under 30% in nonfatal cases (26;44;53).

No case of burn-related mortality with burned TBSA under 30% was reported in Tehran Forensic Medicine Council study (16).

Discussing some methodological issues:

Community-based studies point more precisely than most hospital-based studies to the context in which various kinds of injuries occur. They can take into account the large numbers of injuries that are not seen in hospitals (63).Most burn-related studies in Iran are hospital-based except for a few community–based studies carried out. This presents difficulties for a holistic concept on burn epidemiology over the country. Many Iranian published studies are conducted on a sufficiently large scale, but it seems that the available data collected and the details necessary for prevention work, are not well presented in some of these studies. However, noteworthy in many western studies is the efficient use of available data and prevention-based discussion of information. An example of a very simple small study conducted in the USA 40 years ago is cited as an example here(64). The availability of the current Iranian health network and its high coverage in rural areas is a valuable opportunity that can serve researchers to better define the epidemiology of minor and moderate burns in their geographical area(49). However, maybe due to the lack of an appropriate attitude about research among some health system managers, and also due to the lack of clear-cut regulations regarding health network collaboration with researchers, this may not be very easily accomplished.

The incorrect or inappropriate use of two epidemiological terms namely "Mortality rate" and " Case-fatality rate", are quite obvious in Iranian burn literature. The mortality rate is usually calculated on an annual base (AMR) as follows:

$$AMR = \frac{\text{Number of deaths due to burns (during one year)}}{\text{Population at risk in midyear}}$$

This rate can be calculated for specific groups and can be standardized to enable comparison between different populations. Citing the validity of numerator and denominator values in the fraction calculating burn mortality, constitutes a necessity for hospital-based studies of the mortality rate. Researchers should discuss how sensitive the study has been recording burn related deaths or the numerator, in the population mentioned as denominator. The following questions are also good to be answered in this respect:

Is the study hospital, the only center providing services to the burned patients for the study population?

What fraction of burned patients, in a province of interest, have not been admitted to the hospitals studied, due to referral to other hospitals located in the neighboring provinces due to proximity or owing to the better services offered at these neighboring hospitals? Such issues are especially noticeable in economically less developed provinces of the country.

How will the non-hospital mortalities be incorporated? These include also the deaths directly referred to the forensic medicine services.

Are out of province mortalities recognized and corrected in the numerator?

In case of regional or capital city hospitals that receive referrals from other provinces, are such referrals removed from the numerator when estimating the rates for the main province in which the center is located?

Have self-consent discharges or out of province hospital referrals been considered by the researcher?

Is the source of figures in the denominator and its appropriateness clearly declared and discussed?

Concerning what is discussed above, any burn mortality rate reported by hospital-based studies should be interpreted cautiously. Case-fatality rate can be an easier alternative to achieve and interpret, in hospital-based studies without enough reliable information to estimate the mortality rate. The conversion factor between the fatality rate and the mortality rate is the incidence(65).

Considering the lower survival time for fatal burns compared to chronic diseases maybe it is possible even to ease the terminology a bit and use fatality ratio which is also used in epidemiological terminology(66).

Case-fatality(CFR)can be calculated as below.

$$CFR = \frac{Number\ of\ deathes\ (in\ a\ given\ time)}{Total\ number\ of\ burn\ victims}$$

Obviously, the case fatality in hospital-based studies applies to the hospital or setting from which the numerator and denominator figures have been extracted. Since, many patients with minor burns are not referred to hospital and some others may die before arriving in hospital, the calculated fatality rate may not be generalized to all burns, which have occurred in the population of interest. An appropriate interpretation of the fatality rate will not be possible unless the situation regarding severe burns cases referred to higher-level centers is well clarified.

Most of the published studies have collected post-event (post-injury) data, and adequate attention is usually not paid to environmental, object (appliance) related and human related factors, at the event or pre-event phases of the injury. In other words, most studies are not specifically designed to produce prevention-oriented results. Some of these studies despite claiming to be prospective studies may only be based on the review of medical files in burn centers, without attempting to collect further information regarding prevention or patterns of injuries. Including an open-ended question in medical records to describe injury pattern of occurrence, can be beneficial for future quantitative or qualitative analysis of content. This may also be of help in validity evaluation and improvement.

To assess the prevalence and incidence of burns, especially minor and moderate burns, the community-based studies are quite vital. In addition, community-based studies can also be of great benefit in studying burn injury predictors. Nevertheless, analyzing the causes and risk factors of burns through appropriate designs in hospital-based case-control studies is also possible, more feasible, and less expensive.

Case–control design is a valuable money and time-saving method to study possible predictors of burns or their aftermaths. Unfortunately case-control designs have received little attention from burns researchers in Iran. However some studies are available from other developing countries presenting possible risk factors of burns (67-71). Two case-control studies have been recently published about the risk factors of self-immolation in Iran. However, due to their very small sample size and common source they may suffer from limitations, but can be considered as a good start to study self-immolation risk (72).

Case-crossover design is among the valuable relatively new methodologies. Indeed they are case-control studies. Malcolm McClure first introduced this method in 1991 to assess the effect of transient exposures on sudden outcomes, and then after, it was extended during the following years (73;74). Since, the injuries constitutionally occur in sudden form, such studies may be ideal to assess the contribution of transient exposures in inducing injuries. It is especially important, because controlling abundant confounding factors in studying injuries is hardly possible.

Qualitative studies are of value for assessments in the field of burns, too. Using qualitative studies is relatively new and results can be practiced for all three phases of injury event in burns. Such studies can be much more beneficial when concordantly carried out with quantitative studies. A recent qualitative study in Iran has been published regarding reactions associated with disfiguring burns(75).

Survival analysis is another method which can be used to analyze either burn related mortalities or the efficacy of different therapeutic approaches. Such survival analysis can be practiced in the form of either cohort or clinical trial studies or on preventive interventions as well. This interesting methodology has also not been greatly used by Iranian researchers. However, one study has appropriately used this method in self-immolation research observing higher mortality risk in cases with more than 75% TSBA (RR=2.6, 95% CI: 1.6-4.3), those with burns in the head and neck (RR=2.5, 95% CI: 1.1-5.2), and those with burns in the lower limbs (RR=5.8, 95% CI: 2.2-14.9) (76).

Longitudinal study designs are also of great importance in assessing the epidemiology of injuries and applicable to community interventional projects. In such studies, variables are repeatedly measured and can be acceptably modeled. These studies utilize advanced statistical methods and to the best of our knowledge have not been practiced by Iranian burn researchers.

Using newer statistical methodologies to manage large numbers of variables is another issue of interest that may be recommended to researchers to consider in analyzing risks and patterns in burn injuries(77-79).

3.2 STATISTICAL MODELS

3.2.1 Some concerns in classical regression

Introduction of multivariate regression models revolutionized analysis and interpretation of research data. Confounding control in multivariate regression helped to improve exchangeability in causal inference. Most health science researchers have used or read about the two popular regression models called as linear and logistic regression. These methods are among the most common classical regression models that have many advantages and some limitations as well. Here in I will discuss about some of the concerns regarding these models that can be handled to some extent by supervised statistical methods which are investigated in this thesis.

3.2.1.1 Power and missing values:

The key element in assessing if one has adequate data to fit a particular model involves the number of events per covariate(80). Power and sample size is a critical issue in any statistical test including the multivariate analysis methods. It gets more complicated in the case of multivariate regression models especially as the number of predictors of interest increases. This issue has always been a matter of concern for researchers and sometimes even due to the limitations in estimating power in regression methods, rules of thumb are presented and used by some researchers while some others have also tried to show that these rules of thumb can be relaxed. Most of these simulation studies have been conducted assuming reasonably low numbers of predictors and complete data. Nevertheless the uncertainty in this area is not removed and may be increased as the research has been expanding in recent years to use wider and shorter datasets instead of traditional long narrow data matrices(81-84).

Although a very large number of papers have discussed the missing data in statistical analysis, the most common solution to missing data problems in regression analysis is to delete either cases or variables so that the resulting data set is complete (85).

Many software packages delete partially missing cases by default, and fit regression models to the remaining, complete, cases. This is a reasonable approach as long as the fraction of cases deleted is small enough, and the cause of values being unobserved is unrelated to the relationships under study. Suppose we have 3% random missing value for each of 300 variables in a case-control study. If we try to remove variables or observations, the remaining set for regression analysis will be a small non-representative set, making it very difficult to draw reasonable conclusions. Many methods of imputation are discussed in the literature, however using a modeling technique, which is relatively robust in this regard, may be the best solution in such a situation.

3.2.1.2 Multicollinearity:

Multicollinearity is the fact that a predictor variable is a linear combination of the other regressors. The stronger this dependency, the larger the standard errors for estimators of this predictor variable, leading to widening of confidence intervals and instability of regression coefficients. Multicollinearity undermines the statistical significance of an independent variable(86). Nevertheless, it must be known that this is not a problem of misspecification(87).

A measure of overall multicollinearity of the variables can be obtained by computing the condition number of the correlation matrix. The condition number can be defined by:

$$\kappa = \sqrt{\frac{\text{maximum eigenvalue of the correlation matrix}}{\text{minimum eigenvalue of the correlation matrix}}} = \sqrt{\frac{\lambda_1}{\lambda_p}}$$

A large number indicates evidence of strong collinearity(87).In the case of perfect multicollinearity, each independent variable is an exact linear combination of all the other independent variables. This should result in it being impossible to estimate any of the parameters of the regression model, as the matrix of X . X will not invert. The total absence of multicollinearity occurs when there is no correlation whatsoever between any combinations of the independent variables. However the situation is usually not as in these two extremes(88).

If there is no multicollinearity, the R2j (i.e. the coefficient of determination of the regression of xj on all predictors in the model) will be zero. Then the variance as well as the estimated coefficient will be the same for the total and partial regression coefficients. However the correlation among predictors leads to inflation of coefficients by the quantity $(\frac{1}{1-R_j^2})$. So the variance inflation factor(VIF) can be calculated as a measure of existing multicollinearity. But practically there is no significance test for it and cutoff value will be based on practical considerations. A popular cutoff value is 10 in this regard(89).

3.2.2 Principal component analysis

Principal component analysis (PCA) is an older method suitable for analyzing correlated data. It can be considered as a historical starting point for development of recent modern models. PCA involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on(90).

There are usually two objectives sought by PCA

To discover or to reduce the dimensionality of the data set.

To identify new meaningful underlying variables

Let's assume that the multi-dimensional data have been collected in a data matrix, in which the rows are associated with the cases and the columns with the variables. The mathematical technique used in PCA is called eigen analysis: we solve for the eigenvalues and eigenvectors of a square symmetric matrix with sums of squares and cross products. The eigenvector associated with the largest eigen value has the same direction as the first principal component. The eigenvector associated with the second largest eigen value determines the direction of the second principal component. The sum of the eigen values equals the trace of the square matrix and the maximum number of eigenvectors equals the number of rows (or columns) of this matrix.

From a mathematical point of view the first principal component, y1, is a linear combination of x1, x2, x3, ...,xp

$$y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p = \sum_{i=1}^{p} a_{1i}x_i$$

Such that the variance of y1 is maximised given the constraint that the sum of the squared weighths is equal to one.

$$\sum_{i=1}^{p} a_{1i}^2 = 1$$

The random variables,xi ,can be either derivation from the main scores or standardized scores. If the variance of y1 is maximized, then so is the sum of the squared correlations of y1 with the original variables x1, x2 ,…. Xp. Principal component analysis finds the optimal weight vector(a11 ,a12 ,….a1p) and the associated varience of y1 which is usually denoted by λ1. The second principal component, y2 , involves finding a second weight vector(a21 ,a22 ,….a2p) such that the variance of

$$y_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p = \sum_{i=1}^{p} a_{2i}x_i$$

Is maximised suject to the constraints that it is uncorrelated with the first principal component and

$$\sum_{i=1}^{p} a_{2i}^2 = 1$$

.

This results in y2 having the next largest sum of squared correlations with the original variables. However the sum of squared correlations with the original variables, or equivalently, the variances of principal components get smaller as successive principal components are extracted. This process can be continued until as many components as variables have been calculated(91).

Statistically, PCA finds lines, planes and hyperplanes in the k-dimensional space that approximate the data as well as possible in the least squares sense. It is easy to see that a line or plane that is the least squares approximation of a set of data points makes the variance of the coordinates on the line or plane as large as possible(1)(Figure 1).

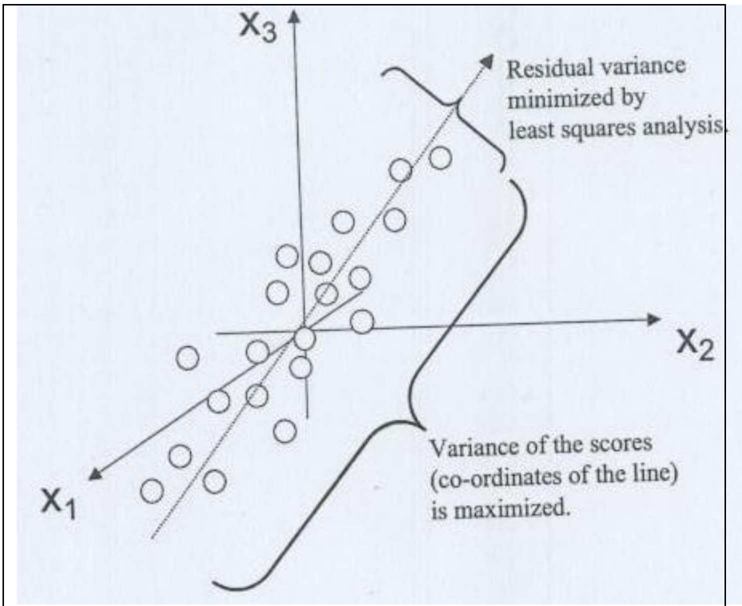


*Figure 1: PCA may be understood as maximizing the variance of projection co-ordinates

When two principal components have been extracted, they together define a plane which is a window into the K-dimensional variable space(Figure 2).

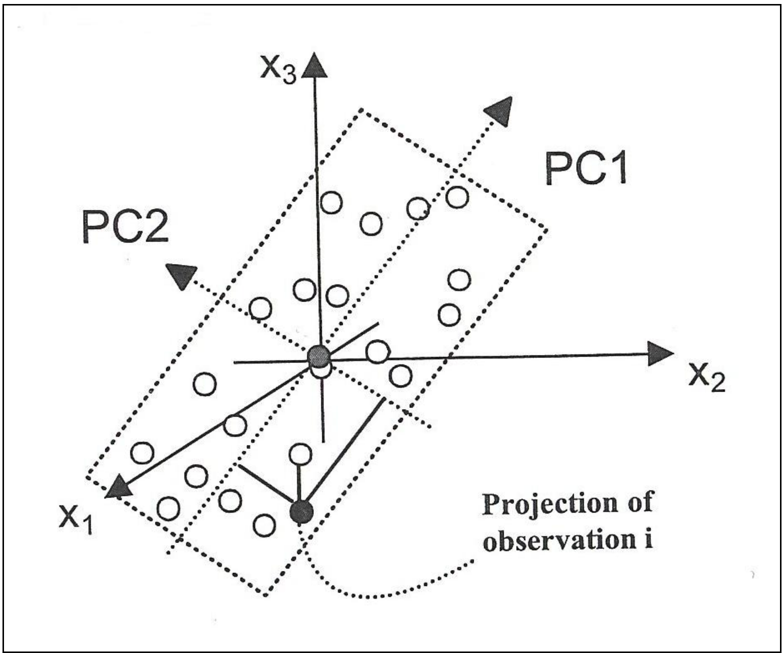


*Figure 2: Two principal components form a plane which is a window into the multidimensional space

By projecting all the observations onto this low-dimensional sub-space and plotting the results, it is possible to visualize the structure of the investigated dataset. The coordinate values of the observations on this plane are called scores. Plotting of such a projected configuration is known as a score plot in component based statistical methods. An example of a score plot is cited from Paper 4 of this thesis project (Figure 3).

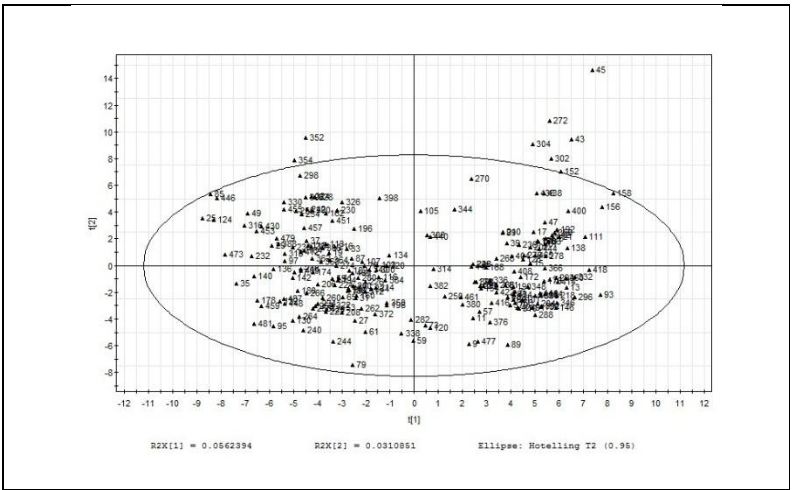


Figure 3 : Score plots in principal component analysis showing existence of two possible clusters of observations

Each observation in this plot is a burn victim hospitalized in Ardabil provincial burn center. The x axis is showing the scores for the first component and y axis shows the scores for the first component.

3.2.3 PLS

Partial least squares is a family of regression based methods designed for the analysis of high dimensional data in a low structure environment(92). It derives its usefulness from its ability to analyze data with many, noisy, collinear, and even incomplete variables in both X and Y(1). The PLS approach was originated in sixties by Herman Wold originally in field of social sciences(economy), later getting popular in chemometrics for the modeling of complicated data sets in terms of chains of matrices, but starting again to become a tool of choice for multivariate analysis in social sciences(93). Herman Wold developed a simple but efficient way to estimate the parameters in these models called NIPALS (Nonlinear Iterative Partial Least Squares)(1).

PLS is a regression extension of PCA, which is used to connect the information in two blocks of variables, X and Y, to each other (Figure 4). PLS provides many model parameters and other residuals-based diagnostic tools, which are useful for understanding and interpreting the acquired regression model.

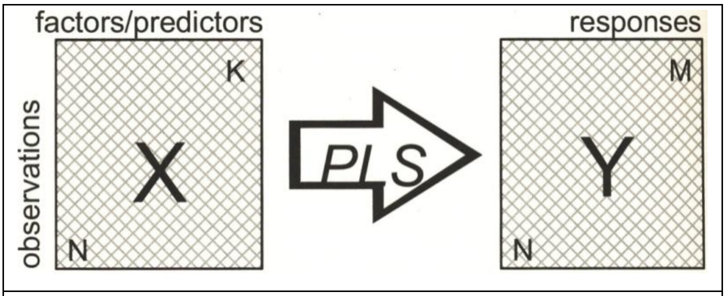


*Figure 4: Schematic view of PLS linking blocks of X and Y variables

PLS can be seen as a particular regression technique for modeling the association between X and Y, but it can be seen as a philosophy of how to deal with complicated and approximate relationships.

In particular, PLS has properties similar to thermodynamic models, in that models linking macroscopic variables are formed without recourse to microscopic theory. This does not mean that PLS is inconsistent with microscopic and detailed theory, but rather that PLS uses the regularities in the data themselves as building blocks in the models, instead of hypothetical functional forms derived from an underlying theory. For this reason, projection models like PLS also work in cases where no good fundamental theory exists, and also provide an independent check on the validity of models derived from theory, where these do exist(1).

The first PLS component is a line in X-space and another line in Y-space (Figure 5). These two lines are calculated in a way to well approximate the point swarms in X and Y. The lines intersect with the average points and by projecting the observations onto them one obtains the scores t1 and u1, for X and Y respectively(1).

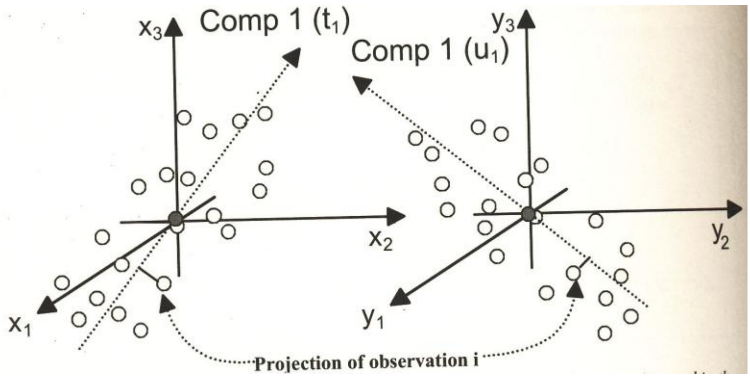


*Figure 5: Interpretation of a PLS component as two lines in X and Y spaces

Usually one PLS component is insufficient to adequately model the variation in the y-data. Similarly as in first component, the second component is also represented by two lines in X and Y spaces. In the X space, the second line is orthogonal to the first one, whereas in the Y space this is not necessary. A two-component PLS model can be interpreted as two planes in X and Y spaces (Figure 6).

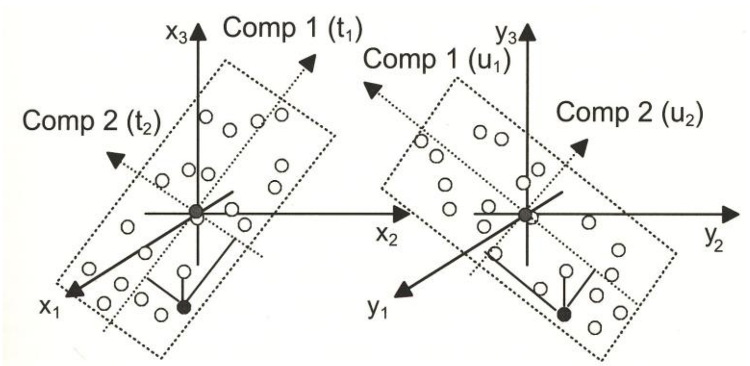


*Figure 6: Model planes comprising the PLS components in X and Y spaces

Historically, the PLS methodology has become popular and proven successful mostly in some application areas like: quantitative structure-activity relationships (QSAR) modeling, multivariate calibration, and process monitoring and optimization. However, later it has found its way into other areas of research also.

PLS and other projection methods have a theoretical foundation based on perturbation theory of a multivariable system. This derivation shows that projection models can approximate any data table as long as there is a certain degree of similarity between the observations (matrix rows). And the approximation is better the greater the similarity between the observations and the greater the number of model components. More on theory of PLS and mathematical algorithms of PLS can be found in references cited above.

Preprocessing of data:

Although here in this thesis I am discussing the preprocessing of data in PLS section, however, data are also pre-processed prior to using other statistical methods including PCA, PLS and OPLS. PLS modeling works best when the data are fairly symmetrically distributed and have a fairly constant "error variance".

In addition, data are usually centered and scaled to unit variance before the analysis. This is because in any given variable will have an influence on the model parameters which increases with the variance of the variable.

A graphical illustration of the impact of mean centering and scaling presented in figure 7 shows that after mean centering and unit variance scaling, all variables will have equal length and mean zero.

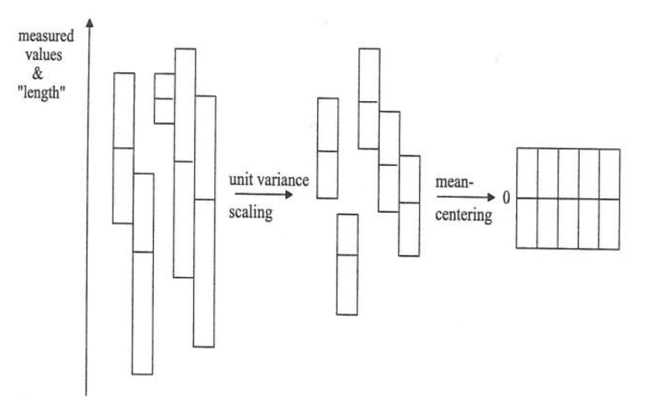


*Figure 7: Graphical illustration of the impact of mean centering and scaling

3.2.4 OPLS

Orthogonal Projections to Latent Structures (OPLS) is a linear regression method that has been employed successfully for prediction modeling in various biological and biochemical applications. OPLS is a modification of the usual PLS method which filters out variation that is not directly related to the response. The result is more transparent models which are easier to interpret(77). The OPLS method was first proposed in 2002 and is a modification of the original NIPALS PLS algorithm. The development of OPLS has, like OSC, been driven by the large amount of non-correlated variation present in data sets today, especially in a multivariate calibration situation. The OPLS method can be seen as either a pure preprocessing method to remove systematic orthogonal variation from a given data set X, or it can be made an integrated part of the regular PLS modeling to provide simpler models and with the additional advantage that the orthogonal variation can be analyzed separately.

OPLS has several advantages over routine partial least square regression modeling and other similar available methods.

1- OPLS method yields more parsimonious PLS models and easier interpretation, because the non-correlated variation and the correlated variation have been separated.

2- OPLS gives an improved detection limit for outliers in the scores, because the non-correlated variation in X could have different statistical distributions than the correlated variation, producing a disturbance in the calculation of e.g. the Hotelling T statistic.

3- An advantage with OPLS compared to the earlier proposed OSC method is that no time-consuming internal iteration is present, making it very fast to calculate(94).

4- Removing the orthogonal principal components from the original data has shown to decrease the total number of components used drastically.

Although OPLS has its roots in multivariate calibration and development of various filters to remove undesired variability in spectral data, but it is not restricted to this field and has general applicability. Omics research like genomics, proteomics and metabolomics are well suited for these modeling techniques. However, in many other fields of research specially medical research where there are large number of variables, these methods can be applied. The general capabilities of these methods poses the possibility of using them also for variable selection in epidemiological research(95). In addition, OPLS is of particular value in spectroscopic calibration and QSAR modeling but is set to become the standard in all areas of data analysis. The mathematical algorithm of OPLS is given in Appendix 2.

3.2.5 PLS-DA & OPLS-DA

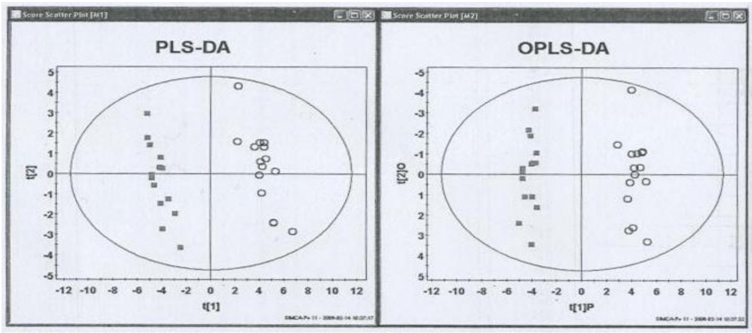


*Figure 8 : A two-class discriminant problem in which OPLS has the predictive power concentrated into the first component, that improves model transparency and interpretability

Discriminant analysis is a well known statistical method used in different areas of research. Classical discriminant analysis methods are compared with logistic regression modeling before(96;97). Later PLS-DA was shown to be quite beneficial for multivariate situations. Considering the advantages of Orthogonal PLS over simple PLS, applying an OPLS discriminant analysis will keep the known advantages of OPLS modeling in field of discriminant analysis(98)(Figure 8).

In some areas of medical research these advantages can be life saving for datasets with large number of correlated possible predictors. For example case control studies or prognostic studies with imaging , proteomics or genomic predictors maybe choice for application of OPLS-DA(99-103).

4 AIMS

The overall aim of this thesis was to analyze some epidemiological patterns and assess the appropriateness of supervised statistical models to investigate burn injury risks and patterns

Specific objectives:

1. Map out epidemiological features of minor and moderate burns using a suitable framework

2. Map out epidemiological features of burn hospitalizations and merging the results of a quantitative content analysis in this regard

3. To assess applicability and validity of the new OPLS-DA statistical model for injury risk assessment, and to compare it with partial least squares regression

4. To investigate the possibility of using supervised statistical models to assess burn injury patterns, outcomes and their interrelationship

5 METHODS

5.1.1 Overall methodology

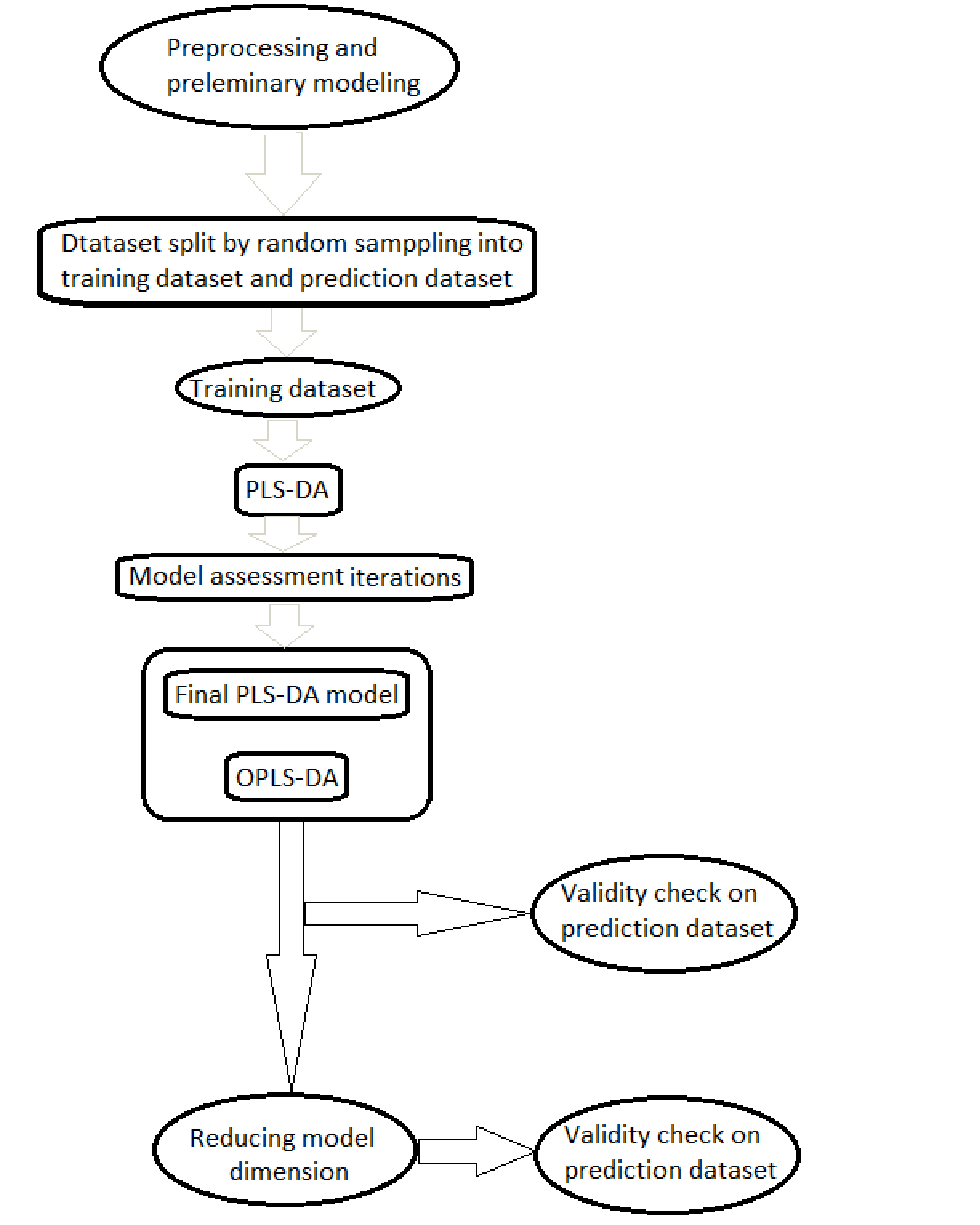
Papers 1, 2 and 4 were prepared based on a joint Swedish-Iranian research project conducted in Ardabil Province, in the north-west of Iran. Paper 3 used data from a case-control study in Tabriz, which is the referral center for several provinces in the north-west of Iran.

Study 1 was a descriptive study conducted on a yearly basis. Participants in Paper 1, were victims of minor and moderate burns living in villages covered by health houses in the rural areas of Ardabil Province. A total of 1700 minor and moderate burn injuries were studied using a pretested questionnaire.

Study 2 was also mainly a descriptive epidemiological study on burn hospitalizations. Participants in Paper 2 were burn victims hospitalized in the Ardabil provincial burn center. Data were collected using the questionnaire in Study 1. Answers to an open-ended question were also coded and analyzed using simple methods. Study 3 was a statistical study assessing and comparing applicability, interpretability and validity of PLS-DA and OPLS-DA statistical models for discriminant analysis in injury epidemiology. Paper 3 used data from a case-control study conducted in Sina burn center in Tabriz. Study 4 was also a statistical study applying PLS-DA and OPLS (for continuous outcome). However, unlike Paper 3, it was not a pure methodological modeling study. Participants in Paper 4 were the same as in Paper 2 except those for whom we lacked complete information to cover the objectives of the study.

5.1.2 Modeling process

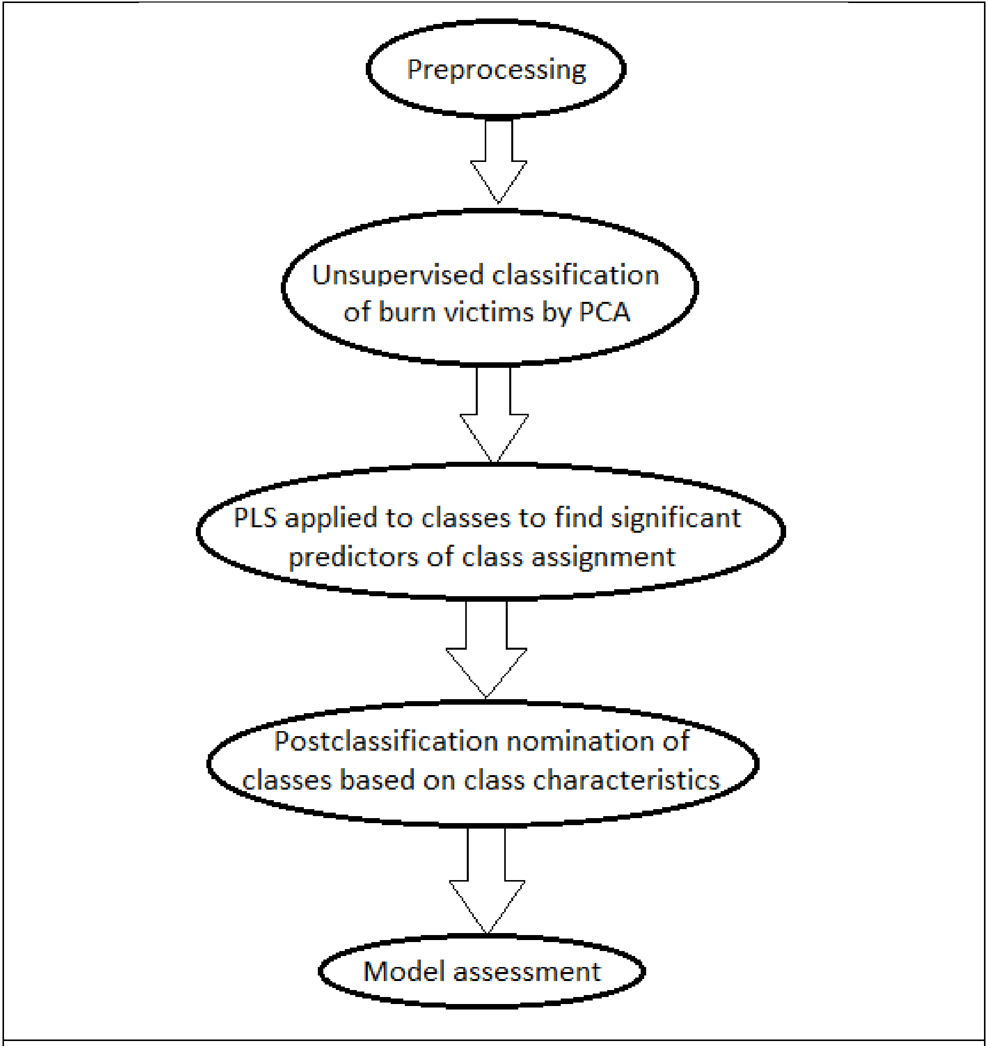
The modeling process in Study 3 is summarized in Algorithm 1:



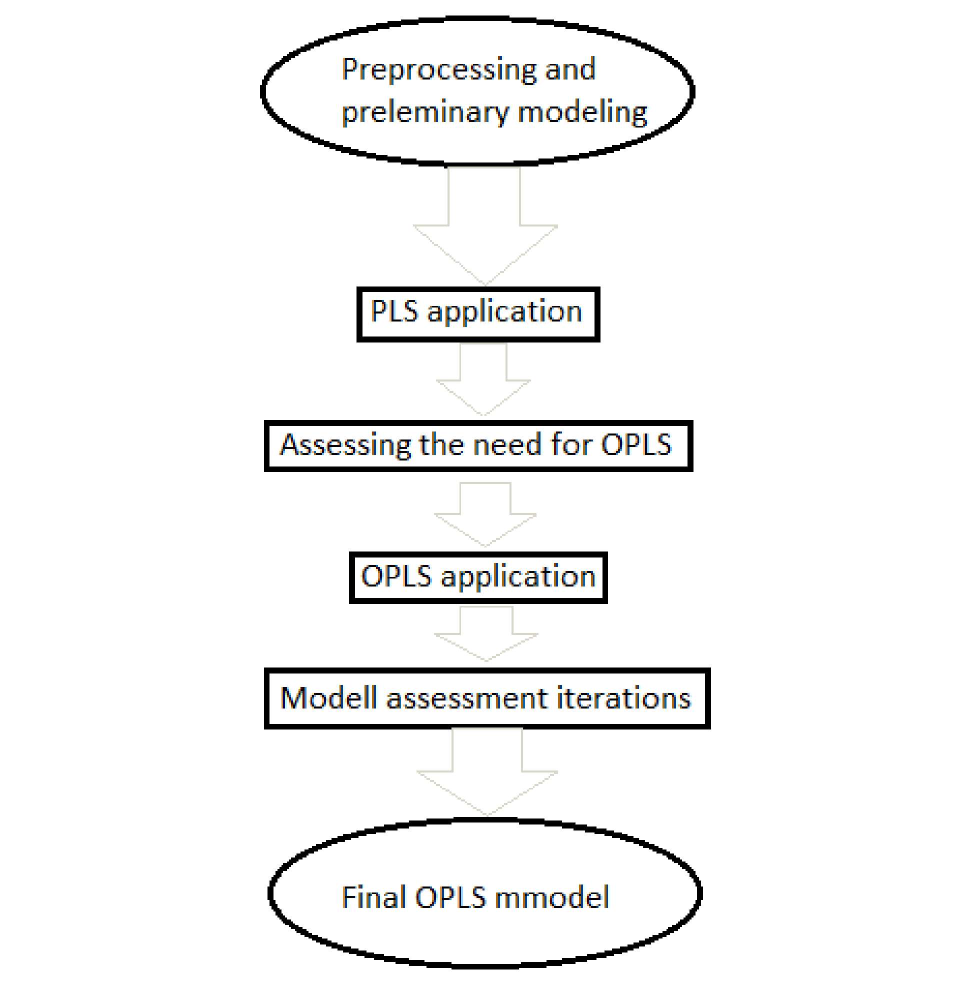
Algorithm 1: Summary of the modeling process in study 3

The modeling process in Study 4 is summarized in Algorithms 2 & 3. Algorithm 2 illustrates the classification modeling process and Algorithm

3 shows the application of the OPLS model in order to predict burn extent measured as total body surface area(TBSA) burned.



Algorithm 2: Summary of the modeling process in classification part of study 4



Algorithm 3: Summary of the modeling process to predict TBSA in study 4

5.1.3 Model assessment and diagnostics

Below are the main model assessment methods used in Paper 3 and Paper 4, with some details on those that are more specific for supervised models.

1- Data were primarily assessed for validity of values with repeated referral to the paper files and in very few case rechecking the medical files.

2- Preprocessing the data by checking for the distributions unit variance scaling and mean centering. Value distributions were reshaped for a few variables using the simplified process in SIMCA P+ software package.

3- Prior to fitting supervised models a preliminary principal component analysis was done for data overview, detecting outliers and groups among the observations.

4- Model goodness fit was assessed using R^2 however as this is an inflationary measure and rapidly approached unity as the model complexity increases, Q^2 was used to assess model predictability. Q^2 is not inflationary and doesn't approach 1 with increased model complexity. Q^2Y also increases with increasing model complexity, but at a certain degree of complexity, predictive ability doesn't improve any further. So combined assessment of $R^2$and Q^2 is kind of cross-validation as a trade-off between fit and predictive ability(figure 9).

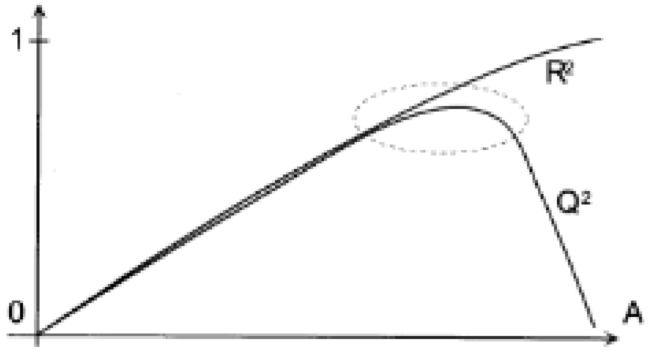


*Figure 9 : Pattern of change in R^2 compared to Q^2, with increasing model complexity(X axis)

5- Response permutation : To provide a measure of statistical significance for the predictive power in cross-validation, response permutation was used. In this process the X-data are left intact while Y-data are permuted to appear in a different order. The model is then fitted to the permuted Y-data and by using cross-validationR^2Y and Q^2Y are computed for the derived model.

6- Leverage was assessed using Hottelings T^2 to which it is proportional.

7- Variable influence was assessed using VIP(variable importance in the projection) measures. It is calculated in SIMCA software package as follows: For a given PLS dimension, a, $(VIN)_{ak}^{2}$ is equal to the squared PLS weight $(w_{ak})^2$ of that term, multiplied by the explained SS of that PLS dimension. The accumulated value ,

$$VIP_{ak} = \sum_a (VIN)_k$$

Was then divided by the total explained SS by the PLS model and multiplied by number of terms in the model. The final VIP will be square root of the number.

$$VIP_{Ak} = \sqrt{\left(\sum_{a=1}^{A}\left(w_{ak}^2 * (SSY_{a-1} - SSY_a)\right) * \frac{K}{(SSY_o - SSY_A)}\right)}$$

8- Critical distance to the model: DModX as observation absolute distance to the model was calculated as:

$$s_i = \sqrt{\frac{\sum e_{ik}^2}{(K-A)}} * v$$

The summation is made over the X variables (k) and e_{ik} are the X-residuals of observation i. "v" is a correction factor (function of the number of observations and the number of components) and is slightly larger than one.

DModY as absolute distance to the model in Y space was calculated as:

$$s_i = \sqrt{\frac{\sum f_{im}^2}{M}}$$

The summation is made over the Y variables (k) and f_{im} are the Y residuals of observation i.

The Normalized distance to the model for X and Y spaces were calculated as follows:

$$S_0 = \sqrt{\frac{\sum\sum e_{ij}^2}{(N-A-A_0)*(K-A)}} \quad \text{in the X space}$$

and

$$S_0 = \sqrt{\frac{\sum\sum f_{im}^2}{(N-A-A_0)*M}} \quad \text{in the Y space}$$

Based on our preprocessing in our study, the A_0 was assumed as 1.

The ratio $(\frac{S_i}{S_0})^2$ which is approximately F distributed was used to compute the critical distance to the model with our preferred probability level.

9- Validation out of data used for modeling: In paper 3 this was done and 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.

10- Model significance testing was done using CV-ANOVA methodlogy(104).

11- Residual distribution graph(Figure 10) and observed vs. predicted graphs were also plotted both through analysis for paper 3 and paper 4(Figure 11).

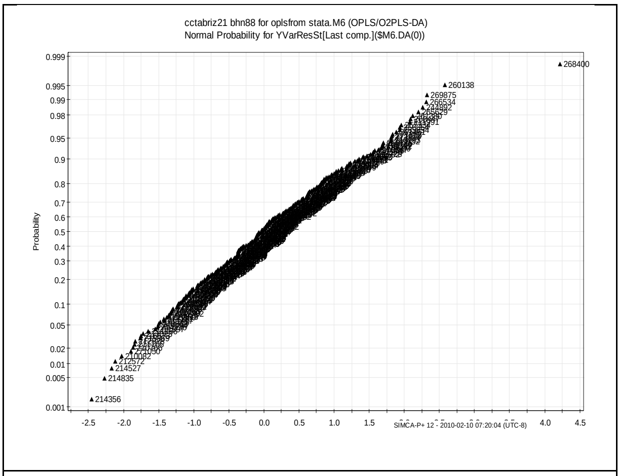


Figure 10 : example of distinguishing an outlier on residual distribution plot

As can be seen in Figure 10, observation number 268400 in one of the fitted models is a residual outlier both regarding the score value and point location.

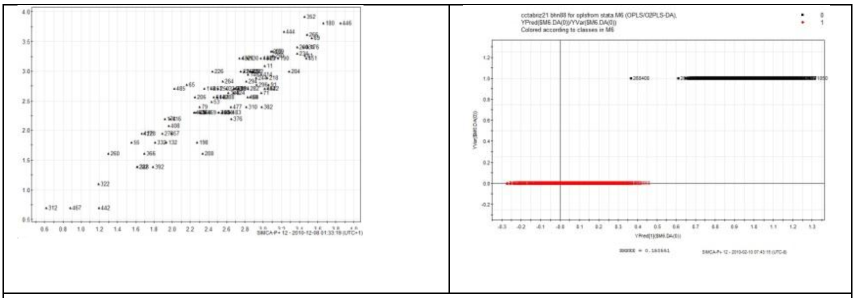


Figure 11 : Observed versus predicted plots assessing model prediction ability

The figure on the right belongs to an OPLS-DA model, like in Paper 3, while the figure on the left is an example of OPLS for a continuous outcome as in Paper 4

12- Although not necessary to be included in Paper 3, receiver operating curves were also plotted to compare PLS-DA and OPLS-DA models in Study 3.

5.1.4 Ethical issues

All four studies had ethical permission from the responsible committees of ethics and were conducted following ethical regulations. The first, second and fourth studies were approved by the Committee of Ethics in Ardabil University of Medical Sciences, and the third study was approved by the Committee of Ethics in Tabriz University of Medical Sciences.

6 RESULTS

Study 1 enrolled 1700 burn injuries, the majority of whom were females and children. Study 2 enrolled 237 burn victims with a slightly higher percentage of males. Some basic results of Study 1 conducted on minor and moderate burns and Study 2 conducted on burns leading hospitalization are compared here in Table 1.

Table 1: Some results compared between study 1 & 2

Variable	Statistic	Study 1	Study 2
Incidence rate	By 100,000 person-years	340	13.2
TBSA	Mean; Median	1.3;≈0	20.5; 15
TBSA, males	Median	-	16.5
TBSA, females	Median	-	15
Age	Mean	19.4	21.6
Age, males	Mean; Median	14.2;7	23.3; 19
Age, females	Mean; Median	23; 18	20.3; 14.5
Gender	(%) Male	41.2	55.7
	(%)Female	51.8	44.3
Place	(%)Home	90.1	81.7
	(%)Work place	9.9	18.3
Place for domestic	(%)Kitchen	37.3	47.1
	(%)Out of kitchen	62.7	52.9
Type of Agent causing burn	(%)fire	5.8	47.3
	(%)Hot liquids & steam	75.1	45
	(%)Other	19.1	7.7
Injury mechanism in scalds	(%)Overturning a container	56	86.3
	(%)Other	44	13.7
Container in scalds	(%)Samovar	7.9	36.5
	(%)Kettles	37.8	31.8
	(%)Pots	13.6	22.4
	(%)Other	40.7	9.3

Study 2 provided more information regarding injury outcome as well as injury patter and descriptive co-variation of them. Bumping into a hot liquid container was the main scenario for a scald injury constituting nearly 70 % of the cases. The difference between flame and non-flame burns in the distribution of burns in extremities was not statistically significant, but head and neck burns were 3.7 times more likely to be caused by flame. The two most important injury patterns, more common among women, were getting burned while using a camping gas stove or while refilling the fuel chamber of kerosene burning appliances without first extinguishing them. 70.1 % of the participants were discharged as improved, 6.1 % died, 11.7 % were referred by ambulance to the highly specialized regional burns center in Tabriz and 12.1 % of them were released from hospital with informed consent.

Application of orthogonal projections to latent structures discriminant analysis(OPLS-DA) in Study 3 to model large numbers of variables in our ongoing burn injury case-control study was found to be successful. The model results and specifications were also compared with discriminant analysis done by partial least squares regression (PLS-DA). Models were found to be promising in applicability both in primary assessments and cross validation done when applied to the training dataset and also after being tested for validity when applied to prediction(Validation) dataset. To discriminate burns cases from controls, it was found that up to 87 variables had regression coefficients significantly different from zero at 95% confidence level for the OPLS model. 107 variables for the first PLS component , 74 variables for the second PLS component and 65 variables for the third PLS component were found to be statistically different from zero. This was the main area where OPLS-DA presented its higher interpretability and reliability and generally both models suited the data well, despite the fact that OPLS-DA was preferable regarding interpretability and even diagnostics.

Application of PLS-DA in Study 4 was successful in defining the characteristics of two different classes of burn victims. Comparing these characteristics convinced the researchers later to call them higher-risk and lower-risk groups when considering both the pre-injury and post-injury related predictors of the classes. Details of variables defining the two classes are given in Table 2. Also in this study we found a noticeable amount of orthogonal variation when trying to predict burn extent by other variables of interest.

Table 2 : Statistically significant predictors of belonging to separate classes modeled by partial least squares regression discriminant analysis

Class A	Class B
Using kerosene, gasoline or coal burning air heaters	Using electric or gas burning air heater
Heating fuel not piped gas	Piped gas being the heating fuel
Cooking fuel not piped gas	Piped gas being the cooking fuel
Not having piped gas facility at home	Having piped gas facility at home
Using valors* for cooking at home	Not using or rare use of valors* for cooking at home
Ceiling made of flammable material	Non-flammable structure of house ceiling
No gas stove safety checkups done in the six months before interview	Gas stove safety checkups done during the last six months
Didn't know the phone number for emergency medical services	Knowing the phone number for emergency medical services
Using kerosene burning samovar* or stove-top samovars and kettles to prepare tea	Referral to a health center
Didn't know the police phone number	Using electric and piped-gas burning samovars, versus other types mentioned in class A to prepare tea
First referral to a health house*	Knowing the phone number to police
Samovar type as: kerosene burning, coal burning or stove-top samovars	Not- Samovar type as kerosene burning, coal burning and stove-top samovar
Always or often cooking in corridor or vestibule	Not always or not often cooking in corridor or vestibule at house
Using non-safe cooking appliances	Not using non-safe cooking appliances
Always using camping gas stove for home cooking	Not always using camping gas stove for home cooking
Not knowing fire services phone number	Knowing the phone number to fire services
Cooking in a non-safe place	Cooking in a safe locations
Lacking a separate kitchen	Male gender
Female gender	Not-using tendir* for cooking bread
Using a tendir* for cooking bread	Having a separate kitchen at house
Always or often cooking in living room	Not always or not often cooking in living room
Never or rarely using classical multi-burner gas stoves	Using classical multi-burner gas stoves for cooking
Use of camping gas stove at home	Not using camping gas stove for cooking at home
Higher total body surface area (TBSA) burned	Using approved samovars with national standard mark
Not using samovars produced under the supervision of national standard organization or not knowing about such a standard	Not living in a house made of made of raw-bricks or mud lumps
Living in a raw-brick-made or mud-made house	Hot objects being responsible for burn injury

Death in hospital due to burn injury	Lower limb not involved in burn injury
Lower limb burnt	Pattern such as bumping into a table or cabinet with a hot container(including samovars on it.
Pattern not a case of bumping into a table or cabinet with a hot container on it.	Valor not involved in burn injury event
Valor involved in a way in burn injury pattern	Higher level of education
Injury mechanism not a result of overturning of a container	Injury mechanism not a case of overturning of a container
A faulty cooking/heating appliance associated with burn injury	No faulty cooking/heating appliance associated with burn injury
Low level of education	Not always using single burner gas stoves for cooking
Always using single burner gas stoves	Unintentional burn injury
Intentional burn injury	Cooking in separate kitchen
Rarely or never cooking in a separate kitchen	Higher roofed living area (per household member)
Upper limb not involved in burn injury	Upper limb involved
Using non-safe appliances for air-heating	Burn due to non-classified agents
Frequent accidents of children overturning a frying pan	Using safer appliances for air-heating
Not using cooking gas stove for heating air	No accidents of children overturning a frying pan
	Stove for heating air
*: see appendix of paper 4 for explanation or pictures	

7 DISCUSSION

Comparing the incidence of minor burns with the hospital-based studies including Study 2 in this research, suggests that frequency of burns is underestimated in burn epidemiology literature due to the paucity of information in this field. Although some attention has shifted in recent years towards outpatient and minor burn injury research, however more research is needed to define the epidemiology of minor and moderate burns worldwide (31;105-110). Thanks to the available Iranian health network system, we managed to assess the epidemiology of minor burn injuries in a cost-effective manner, while due to methodological and budgeting limitations it was not possible to define it for urban population of Ardabil Province. We recommend that other researchers in Iran take advantage of such availability or that researchers in other countries that that have access to a similar system, do likewise. The Iranian health network has proved to be quite successful in improving the rural health indices but it can also be considered as an opportunity to expand health research and education. It has also to be added that, there has een a collaboration with the Mississippi health authorities to use the Iranian health house model in USA(111-113). It has been shown that minor burns once comprised 95% of all burn referrals in the USA(110). Gross estimation of the incidence of minor and moderate burns in our study in rural areas of Ardabil Province showed the figure (340 by 100,000 person years) to be more than 25 times the incidence of burn hospitalization in Ardabil provincial burn center. In spite of very slight denominator uncertainty, the main concern of incidence underestimation may be on numerator in the second study. Comparing the results of burns in Paper 1 and Paper 2 provides clues regarding the differences in patterns of minor vs. severe burns but caution should be taken due to the inclusion of urban victims in Paper 2.

Gender and age were studied in more detail in this thesis especially in Paper 1 and Paper 4. There is a strong argument for defining adult women and children under the age of seven of both genders, as the priority target groups for possible burn prevention plans in Ardabil Province or similar settings. Teenager girls are not as yet the focus of most unintentional burn injury research, However, it has also been suggested that they too are an important target group, because in a cultural setting like rural and even urban Iran, they are expected to learn cooking for their future role as a housewife. Compared to adult women, if adjusted for the exposure, they may have higher risk of being burned due to their lack of competence and experience in this area (114). Gender vulnerability trend reversal over age, seems to be a quite convincing finding at least for minor burns. This has also been reported in few reports earlier (115;116).

Another field of importance in our results concerned the cooking and heating appliances. Most Iranian studies in this regard have not considered measuring specifically regarding this issue. This has been focus of interest in some studies from LMICs, but mostly in the reporting of descriptive statistics leading even to some ignorance in epidemiological reviews(14;117-119).

Although a bit late in doing so, burn researchers have also started to use content analysis from 1990s (75;120-128). All of these studies have used qualitative content analysis, and to the best of our knowledge, the use of quantitative content analysis has not been reported prior to this work. Even in this thesis, it was not the focus of study but it seems that the application of supervised models can pave the way for less complicated analysis of possibly correlated data coming from this type of research.

There is a high variety of possible predictors of burn injuries. These predictors can be categorized as human related predictors, environmental predictors and object (mainly heating-cooking appliance) related predictors (14;129;130). This usually leads to abundant numbers of variables being studied in injury research. In Study 3 which was a case-control study, we used up to 250 variables in models and similarly in Study 4 we had a total of 147 variables. This is while we had not measured all possible variables. For example, we have not measured much on the clothing situation or about the time between the injury occurrence and arrival at a hospital. In addition, interactions are not evaluated in this project. A similar situation regarding the number of variables is expected in other fields of injury epidemiology. For example, regarding fall injuries, as many as 400 variables have been suggested to be of interest(131). Considering this large number of variables and possible complex or pair-wise correlations among these variables, use of latent variable based statistical methods can be considered. However, as in this thesis, researchers are usually interested to in making predictions and studying the relation between X and Y variables as well as the internal correlations among each type of variable. Therefore, the first generation of supervised modeling techniques like PLS or PLS-DA can be of help in this regard. Both in Study 3 and in Study 4 we found these models to be applicable and beneficial. However PLS also suffers some limitations like interpretability and lower capability in the case of existing orthogonal variations. Some later extensions like orthogonal signal correction(OSC) and projected orthogonal signal correction(POSC) were provided to manage these limitations, but they also suffered other pitfalls that limited their use. However the recently presented OPLS and OPLS-DA models seem to be suitable options to overcome some of these limitations(2;77;94;98;131). Nevertheless, as we found in Study 4 about the class predictors, when the numbers of components are fewer, there may be no need to use OPLS as an alternative for PLS. Both in Study 3 and in the prediction of TBSA in Study 4 we found that orthogonal variation existed in the modeling process. This can be considered as an argument for using OPLS models. In this thesis, we used supervised models as alternatives to classical regression models. But they can also be used as a data reduction tool before applying classical regression models, to ease variable selection as a better alternative for available methods like stepwise variable selection. Another field of applicability can be the generation of injury risk scales. Based on the results of fitted models, model diagnostics and model validity assessment results, it seems quite likely that supervised statistical models can be used in risk and pattern analysis in burn injuries, and possibly other fields of injury or public health epidemiology. No doubt, this is conditional on approved consistency of our findings on future research and detailed investigation of other aspects of modeling methodology not investigated here. In addition, in this project, we did not aim to find replacements for classical regression models, that have many advantages over the methods we applied in this work. Instead, we have tried to investigate newer options in the selection of statistical methods available, so that the researcher can choose the most relevant methodology considering the research question or characteristics and scale of collected data.

8 LIMITATIONS:

Due to the limitations in time, budget and methodology, a complete mapping out of burn epidemiology covering both urban and rural areas of the Ardabil province population couldn't be expected in this thesis project. As a rule, researcher errors in design and conduction of the studies should also be added to the above. In addition, insufficiency in interpreting and reporting the research findings can be observed throughout the work. For example in Paper 1, we tried to provide further details about the situation regarding health facilities in Ardabil Province. However, later we found that by adding this extra information, some ambiguity is introduced into defining the study base. Also in Paper 4, due to journal limitations in word count we omitted details of our methods in model assessment, while it was reasonable to split the information into two papers, one discussing the classification of burn victims and one regarding predictors of burn extent. Another limitation was that adding the residential status in classification model in Paper 4, could improve the results even more. However, we did not add a question on this to the interview because we wanted to save time and planned to extract this information from the patient files. Later on however we found that, some patients had given the address of their relatives living in Ardabil where the burn center is located. Therefore, we did not rely on this information and omitted it from the analysis. During the modeling process our aim was to assess model behavior for this type of data rather than model results, so we arbitrarily made our choices among model construction methods to produce acceptable results rather than constructing much more complex but more accurate models. For example we didn't do exhaustive cycles of adding interaction complexes, trying different scaling methods, testing many transformations to correct for any trivial skewness in data and testing different missing value thresholds to exclude variables. As regards the consistency of our findings regarding the appropriateness of these models over repeated research, such issues can be considered as valuable areas for future research.

9 CONCLUSIONS:

This thesis work was focused on some methods and statistical aspects of injury epidemiology that have not been examined by other researchers in Iran or similar settings. The first study explored the acceptable sensitivity of using the rural health network as a facility to study the incidence and epidemiology of minor burns. In spite of a high incidence measured, an incidence underestimation should also be considered. The second study used the same data collection tool in a hospital setting, however took advantage of quantitative content analysis and explored the richness of available information by means of a simple open-ended question. It seemed that some information derived was likely to be of importance for prevention work.

Study 3 and Study 4 were the main methodological studies, and provided compromising information regarding the applicability of supervised statistical models in assessing risks and patterns in injury epidemiology. Based on the results of these studies and considering the potential capabilities of supervised statistical models, from a prevention focus of interest, 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.

To the best of our knowledge, the new OPLS statistical model has not been reported to be used in public health epidemiology since the time it was developed for chemometrics in 2002 and its later extensions in 2006. However, considering the possible methodological limitations in our study and the insufficiency of information on this model, our findings can just be a starting point for detailed future research in this field. The thesis introduces possible applicability of supervised models in injury epidemiology and does not question many advantages of the classical regression models that remain as gold standard in many injury research situations .

10 ACKNOWLEDGEMENTS

I take the first chance here, to heartily express my gratitude to Dr. Reza Mohammadi who was much more than a supervisor for me. I also thank Professor Bangdiwala who supervised my research and guided me in establishing a necessary background, learning about advanced statistical methods and also supervising me to start teaching statistics. I am grateful to Professor Leif Svanström and Associate Professor Robert Ekman who motivated me to dive into the vast field of injury research and guided me through that.

I am very thankful to my halftime seminar committee members, Professor Ilona Koupil, Associate Professor Tahereh Moradi and Dr. Viveca Björnhagen for their review of my papers and their valuable scientific comments on them.

I am thankful to all the faculty at the Department of public health sciences at Karolinska Institutet, specially Professor Lucie Laflamme, Associate Professor Marie Hasselberg, Professor Per Tillgren, Professor Peter Allebeck, Professor Joaquim Soares, Professor Lars-Gunnar Horte, Associate Professor Stephen Lawoko, Dr. Koustuv Dalal and also to the kind staff at our division Lily Mogess, Moa Sunström, Marita Larsson and Anders.

My thanks to all the Professors and lecturers from whom I learnt so much during the courses I took at Karolinska Institutet.

Thanks to my student friends in Karolinska who tried to help me, specially to Gholamreza, Daryoush, Davood, Omid, David, Andres, Hassan, Ali, Ime, Maria and Samia.

My thanks to Dr. Karin Vågstrand who helped in my thesis application and also to other staff at the Karolinska Institutet. My special thanks the staff in the KI library, whose names I regrettably do not remember. They were really kind and supportive to the students.

My thanks to Gary Watson and Teresa DeAngelis for their excellent help in language editing of my thesis.

My heartiest thanks to Associate Professor Shahnam Arshi who was the key person supporting my research in Iran. He is also someone I have always admired regarding his kind behavior to everyone even his opponents.

I thank those who helped in giving me the opportunity to do research in Iran or backed me in my duties while I was studying in Sweden. I am just giving the names of a few among many. Professeor Malekafzali, Professor Kazem Mohammad(From whom I learned a lot as my earlier supervisor), Professor Majidpour, Professor Gooya, Associate Professors, Meshkini, Farhoudi, Savadi, Habibzadeh, Mirzarahimi, Jabbari, and Dastgiri. There were also other people who helped me a lot during my research in Ardabil specially Mr.Rouhi, Mohammadhosseini, Lotfollahzadeh, Dr. Mashoufi, Dr.Hekmat, Dr.Sepehram and Mrs Malekpour.

Mostafa and Morteza are the two very close friends who helped me a lot in Tabriz and I really don't know how to thank them. Also I am grateful to my kind colleagues in RDCC. I thank twice Mrs. Salehnia and Yasin who helped in designing my thesis cover picture.

Finally, I declare my heartiest thanks to my wife Louiz, and my brother Arshad who really and heartily backed me to continue my studies. I thank also my parents and my In-laws, my sisters and my other brother who supported me.

And last and certainly not least my children Yasin and Yasamin, who gave me a non-informed consent to be partly absent from their life while they needed me.

11 REFERENCES

(1) Eriksson L, Johansson E, Wold N, Trygg J, Wikstrom C, Wold S. Multi- and Megavariate data analysis: Basic principals and applications. 1st ed. Umea: Umetrics AB; 2006.

(2) Eriksson L, Johansson E, Wold N, Trygg J, Wikstrom C, Wold S. Multi- and Megavariate data analysis: Advanced applications and method extensions. 1st ed. Umea: Umetrics AB; 2006.

(3) Baker SP, O'Neill B, Ginsburg MJ, Li G. The injury fact book. 2nd ed. New York: Oxford University Press; 1992.

(4) Barss P, Smith GS, Mohan D. Injury prevention: an international perspective. New York: Oxford University Press; 1998.

(5) Aili-low J. It's not just a burn: physical and psycholoogical problems after burns Uppsala Universitet; 2007.

(6) McKibben JB, Ekselius L, Girasek DC, Gould NF, Holzer C, Rosenberg M, et al. Epidemiology of burn injuries II: psychiatric and behavioural perspectives. Int Rev Psychiatry 2009 Dec;21(6):512-21.

(7) Liao CC, Rossignol AM. Landmarks in burn prevention. Burns 2000 Aug;26(5):422-34.

(8) Linares AZ, Linares HA. Burn prevention: the need for a comprehensive approach. Burns 1990 Aug;16(4):281-5.

(9) Mock C, Quansah R, Krishnan R, rreola-Risa C, Rivara F. Strengthening the prevention and care of injuries worldwide. Lancet 2004 Jun 26;363(9427):2172-9.

(10) Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. Lancet 1997 May 17;349(9063):1436-42.

(11) Murray CJ, Lopez AD. The global burden od disease. World Health Organization; 1996.

(12) National burn care review: a review of burn care in The British Islets. http://www.nbcg.nhs.uk/national-burn-care-review/ . 2001.
Ref Type: Internet Communication

(13) World Health Organization. Injury: A leading cause of the global burden of disease, 2000. Geneva: World Health Organization; 2002.

(14) Forjuoh SN. Burns in low- and middle-income countries: a review of available literature on descriptive epidemiology, risk factors, treatment, and prevention. Burns 2006 Aug;32(5):529-37.

(15) Naghavi M, Abolhassani F, Pourmalek F, Jafari N, Moradi M, Eshrati B, et al. Burden of dieases and injuries in Iran. Iranian journal of Epidemiology 2009;4(1):1-19.

(16) Sheykhazadi A, Garadaghi J, Ghadyani MH. Epidemiology of mortal burn injuries in Tehran. Journal of forensic medicine 2005;12(3):151-7.

(17) Akbari ME, Naghavi M, Soori H. Epidemiology of deaths from injuries in the Islamic Republic of Iran. East Mediterr Health J 2006 May;12(3-4):382-90.

(18) Ahmadi A, Mohammadi R, Stavrinos D, Almasi A, Schwebel DC. Self-immolation in Iran. J Burn Care Res 2008 May;29(3):451-60.

(19) Soori H, Akbari ME, Ainy E, Zali AR, Naghavi M, Shiva N. Epidemiological Pattern of Non-Fatal Injuries in Iran. Pakistan Journal of Medical Sciences 2010;26(1):206-11.
(20) Arshi S, Sadeghi-Bazargani H, Mohammadi R, Ekman R, Hudson D, Djafarzadeh H, et al. Prevention oriented epidemiologic study of accidental burns in rural areas of Ardabil, Iran. Burns 2006 May;32(3):366-71.
(21) Lari AR, Alaghehbandan R, Nikui R. Epidemiological study of 3341 burns patients during three years in Tehran, Iran. Burns 2000 Feb;26(1):49-53.
(22) Maghsoudi H, Garadagi A, Jafary GA, Azarmir G, Aali N, Karimian B, et al. Women victims of self-inflicted burns in Tabriz, Iran. Burns 2004 May;30(3):217-20.
(23) Maghsoudi H, Samnia R, Garadaghi A, Kianvar H. Burns in pregnancy. Burns 2006 Mar;32(2):246-50.
(24) Maghsoudi H, Gabraely N. Epidemiology and outcome of 121 cases of chemical burn in East Azarbaijan province, Iran. Injury 2008 Sep;39(9):1042-6.
(25) Mohammadi AA, Amini M, Mehrabani D, Kiani Z, Seddigh A. A survey on 30 months electrical burns in Shiraz University of Medical Sciences Burn Hospital. Burns 2008 Feb;34(1):111-3.
(26) nsari-Lari M, Askarian M. Epidemiology of burns presenting to an emergency department in Shiraz, South Iran. Burns 2003 Sep;29(6):579-81.
(27) Panjeshahin MR, Lari AR, Talei A, Shamsnia J, Alaghehbandan R. Epidemiology and mortality of burns in the South West of Iran. Burns 2001 May;27(3):219-26.
(28) Rajabian MH, Aghaei S, Fouladi V. Analysis of survival and hospitalization time for 2057 burn patients in Shiraz, southwestern Iran. Med Sci Monit 2007 Aug;13(8):CR353-CR355.
(29) Rastegar LA, Alaghehbandan R. Epidemiological study of self-inflicted burns in Tehran, Iran. J Burn Care Rehabil 2003 Jan;24(1):15-20.
(30) Saadat M. Epidemiology and mortality of hospitalized burn patients in Kohkiluye va Boyerahmad province (Iran): 2002-2004. Burns 2005 May;31(3):306-9.
(31) Taghavi M, Rasouli MR, Boddouhi N, Zarei MR, Khaji A, Abdollahi M. Epidemiology of outpatient burns in Tehran: an analysis of 4813 cases. Burns 2010 Feb;36(1):109-13.
(32) Hasani L, Aghamolei T, Boushehri E, Sabili A. Analysis of hospitalized burned patients in Bandar Abbass, Iran. J Res Health Sci 2009;9(1):50-3.
(33) Maghsoudi H, Adyani Y, Ahmadian N. Electrical and lightning injuries. J Burn Care Res 2007 Mar;28(2):255-61.
(34) Sheikhazadi A, Kiani M, Ghadyani MH. Electrocution-related mortality: a survey of 295 deaths in Tehran, Iran between 2002 and 2006. Am J Forensic Med Pathol 2010;31(1):42-5.
(35) Mohammadi AA, Danesh N, Sabet B, Amini M, Jalaeian H. Self-inflicted burn injuries in southwest Iran. J Burn Care Res 2008 Sep;29(5):778-83.
(36) Lari AR, Joghataei MT, Adli YR, Zadeh YA, Alaghehbandan R. Epidemiology of suicide by burns in the province of Isfahan, Iran. J Burn Care Res 2007 Mar;28(2):307-11.

(37) Lari AR, Alaghehbandan R, Panjeshahin MR, Joghataei MT. Suicidal behavior by burns in the province of Fars, Iran. Crisis 2009;30(2):98-101.
(38) Groohi B, Rossignol AM, Barrero SP, Alaghehbandan R. Suicidal behavior by burns among adolescents in Kurdistan, Iran: a social tragedy. Crisis 2006;27(1):16-21.
(39) Ahmadi A. Suicide by self-immolation: comprehensive overview, experiences and suggestions. J Burn Care Res 2007 Jan;28(1):30-41.
(40) Zarghami M, Khalilian A. Deliberate self-burning in Mazandaran, Iran. Burns 2002 Mar;28(2):115-9.
(41) Mohammadi AA, Danesh N, Sabet B, Jalaeian H, Mohammadi MK. Self-burning: A common and tragic way of suicide in fars province, Iran. Iran J Med Sci 2008;33(2):110-3.
(42) Ghaderi R, Attaraan A. A review of burn cases seen in imam Reza teaching hospital-Birjand. JOURNAL OF BIRJAND UNIVERSITY OF MEDICAL SCIENCES 2003;10(1):15-9.
(43) Groohi B, Alaghehbandan R, Lari AR. Analysis of 1089 burn patients in province of Kurdistan, Iran. Burns 2002 Sep;28(6):569-74.
(44) Koushyar H, Amouzgar M, Shakeri M. The epidemiology of burns in Mashhad Imam Reza burn center(MIRBS)[In Persian]. Ofogheh Danesh 2004;10(2):43-50.
(45) Maghsoudi H, Samnia N. Etiology and outcome of pediatric burns in Tabriz, Iran. Burns 2005 Sep;31(6):721-5.
(46) Hosseini RS, Askarian M, Assadian O. Epidemiology of hospitalized female burns patients in a burn centre in Shiraz. East Mediterr Health J 2007 Jan;13(1):113-8.
(47) Aghakhani N, Rahbar N, Feizi A. Epidemiology of burn injuries in West Azerbaijan Province, Western Iran. Iran Red Crescent Med J 2009;11(1):85-9.
(48) Torabian S, Saba MS. Epidemiology of paediatric burn injuries in Hamadan, Iran. Burns 2009 Dec;35(8):1147-51.
(49) Sadeghi-Bazargani H, Mohammadi R, Svanstrom L, Ekman R, Arshi S, Hekmat S, et al. Epidemiology of minor and moderate burns in rural Ardabil, Iran. Burns 2010 Sep;36(6):933-7.
(50) Lari AR, Panjeshahin MR, Talei AR, Rossignol AM, Alaghehbandan R. Epidemiology of childhood burn injuries in Fars province, Iran. J Burn Care Rehabil 2002 Jan;23(1):39-45.
(51) Sadeghi-Bazargani H, Arshi S, Ekman R, Mohammadi R. Prevention oriented epidemiology of burn injuries in Ardabil provincial burn center, Iran. DOI:10.1016/j.burns.2010.09.013. Burns. In press 2010.
(52) Naghavi M, Djafari N, Alaeddini F, Akbari I. Injury epidemiology in the islamic republic of Iran. First ed. Tehran: Iranian MOH Publications; 2004.
(53) Rouzbehani R, Omranifard M, Zamani A, Rouzbehani A, Farajzadegan Z, Rezaei F. An epidemiological study on burned patients admitted to Imam Mousa Kazem hospital, Isfahan, 2003-2004. Sharekord University of Medical Sciences Journal 2005;7(1).
(54) Arshi S, Sadeghi-Bazargani H, Mohammadi R, Mohammad Zadeh MS, Rouhi A, Barak M. Burns comprising 3/4th of home injuries in pre-school children of the rural areas of Ardabil Province, Iran. J Med Sci (Pakistan) 2007;7(2):248-51.

(55) Saadat M, Bahaoddini A, Mohabatkar H, Noemani K. High incidence of suicide by burning in Masjid-i-Sulaiman (southwest of Iran), a polluted area with natural sour gas leakage. Burns 2004 Dec;30(8):829-32.
(56) Alaghehbandan R, Lari AR, Joghataei MT, Islami A. The Role of Marital Status, Literacy, and Urbanity in Suicidal Behavior by Burns in the Province of Khorasan, Iran. Community Ment Health J 2010;1-5.
(57) Taghaddosinejad F, Sheikhazadi A, Behnoush B, Reshadati J, Anary SHS. A survey of suicide by burning in tehran, Iran. Acta Med Iran 2010;48(4):266-72.
(58) Mohammadi R, Ekman R, Svanstrom L, Gooya MM. Unintentional home-related injuries in the Islamic Republic of Iran: findings from the first year of a national programme. Public Health 2005 Oct;119(10):919-24.
(59) Alaghehbandan R, MacKay RA, Rastegar LA. Pediatric burn injuries in Tehran, Iran. Burns 2001 Mar;27(2):115-8.
(60) Karimi H, Momeni M, Momeni M, Rahbar H. Burn injuries during pregnancy in Iran. Int J Gynaecol Obstet 2009 Feb;104(2):132-4.
(61) Maghsoudi H, Aghamohammadzadeh N, Khalili N. Burns in diabetic patients. Int J Diabetes Dev Ctries 2008 Jan;28(1):19-25.
(62) Maghsoudi H, Pourzand A, Azarmir G. Etiology and outcome of burns in Tabriz, Iran. An analysis of 2963 cases. Scand J Surg 2005;94(1):77-81.
(63) Bartlett SN. The problem of children's injuries in low-income countries: a review. Health Policy Plan 2002 Mar;17(1):1-13.
(64) Smith EI. The epidemiology of burns. The cause and control of burns in children. Pediatrics 1969 Nov;44(5):Suppl-7.
(65) Edler L, Kitsos C. Recent advances in quantitative methods in cancer and human health risk assessment. Wiley & Sons; 2005.
(66) James Eeal. Epidemiology, biostatistics and preventive medicine . 3rd ed. Saunders; 2007.
(67) Forjuoh SN, Guyer B, Strobino DM, Keyl PM, ener-West M, Smith GS. Risk factors for childhood burns: a case-control study of Ghanaian children. J Epidemiol Community Health 1995 Apr;49(2):189-93.
(68) Horner BM, Ahmadi H, Mulholland R, Myers SR, Catalan J. Case-controlled study of patients with self-inflicted burns. Burns 2005 Jun;31(4):471-5.
(69) Mashreky SR, Rahman A, Khan TF, Svanstrom L, Rahman F. Determinants of childhood burns in rural Bangladesh: A nested case-control study. Health Policy 2010 Mar 2.
(70) Petridou E, Trichopoulos D, Mera E, Papadatos Y, Papazoglou K, Marantos A, et al. Risk factors for childhood burn injuries: a case-control study from Greece. Burns 1998 Mar;24(2):123-8.
(71) van Rijn OJ, Bouter LM, Kester AD, Knipschild PG, Meertens RM. Aetiology of burn injuries among children aged 0-4 years: results of a case-control study. Burns 1991 Jun;17(3):213-9.
(72) Ahmadi A, Mohammadi R, Schwebel DC, Yeganeh N, Soroush A, Bazargan-Hejazi S. Familial risk factors for self-immolation: A case-control study. J Women's Health 2009;18(7):1025-31.
(73) Maclure M, Mittleman MA. Should we use a case-crossover design? Annu Rev Public Health 2000;21:193-221.

(74) Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. Am J Epidemiol 1991 Jan 15;133(2):144-53.
(75) Rahzani K, Taleghani F, Nikbakht NA. Disfiguring burns and the experienced reactions in Iran: consequences and strategies--a qualitative study. Burns 2009 Sep;35(6):875-81.
(76) Dastgiri S, Kalankesh LR, Pourafkary N, Vahidi RG, Mahmoodzadeh F. Incidence, survival pattern and prognosis of self-immolation: A case study in Iran. J Public Health (Germany) 2006;14(1):2-6.
(77) Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). J Chemometr 2002;16(3):119-28.
(78) Sadeghi-Bazargani H, Shrikant B, Mohammadi R, Mohammad K. Application of the new OPLS-DA statistical modeling technique to manage large number of variables in a burn injury case control study. 3rd meeting of the EURO Working Group on Stochastic Modelling. Naflpio 2010.
(79) Eriksson S, Lundquist A, Gustafson Y, Lundin-Olsson L. Comparison of three statistical methods for analysis of fall predictors in people with dementia: negative binomial regression (NBR), regression tree (RT), and partial least squares regression (PLSR). Arch Gerontol Geriatr 2009 Nov;49(3):383-9.
(80) Hosmer D, Lemshow S. Applied logistic regression. NY: Wiley Interscience; 2000.
(81) Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol 2007 Mar 15;165(6):710-8.
(82) Vittinghoff E, Sen S, McCulloch CE. Sample size calculations for evaluating mediation. Stat Med 2009 Feb 15;28(4):541-57.
(83) Zolotor AJ, Burchinal M, Skinner D, Rosenthal M. Maternal psychological adjustment and knowledge of infant development as predictors of home safety practices in rural low-income communities. Pediatrics 2008 Jun;121(6):e1668-e1675.
(84) Dupont WD, Plummer WD, Jr. Power and sample size calculations for studies involving linear regression. Control Clin Trials 1998 Dec;19(6):589-601.
(85) Weisberg S. Applied Linear Regression. Newjersey: Wiley & Sons; 2005.
(86) Allen M. Understanding regression analysis. NY: Plenum publishing corporation; 1997.
(87) Chatterjee S, Hadi A. Regression analysis by example. 4 ed. Newjersey: Wiley interscience; 2006.
(88) Cameron S. Multicollinearity: serious worry or minor nuisance. Econometrics. McGraw Hill Education; 2005.
(89) Freund R, Wilson W, Sa P. Multicollinearity. Regression analysis: Statistical modeling of a response variable.Burlington: Elsevier Inc; 2006.
(90) Jolliffe I. Principal Component Analysis. 2nd ed. NY: Springer; 2002.
(91) Dunteman GH. Principal Components Analysis. NY: Sage Publications; 1989.
(92) Esposito Visni V, Chin WW, Henseler J, Wang H. Handbook of partial least squares. Springer; 2010.
(93) Abdi H. Partial least squares(PLS) regression. In: Levis-Beck M, Bryman A, Futting T, editors. Encyclopedia of social sciences research methods.Thousand Oaks: Sage; 2003.

(94) Fearn T. On orthogonal signal correction. Chemometr Intell Lab Syst 2000;50(1):47-52.

(95) Walter S, Tiemeier H. Variable selection: current practice in epidemiological studies. Eur J Epidemiol 2009;24(12):733-6.

(96) Manel S, Dias J, Ormerod S. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions. Ecological Modelling 1999;120(2-3):337-47.

(97) Hossain M, Wright S, Petersen LA. Comparing performance of multinomial logistic regression and discriminant analysis for monitoring access to care for acute myocardial infarction. J Clin Epidemiol 2002 Apr;55(4):400-6.

(98) Bylesj M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. OPLS discriminant analysis: Combining the strengths of PLS-DA and SIMCA classification. J Chemometr 2006;20(8-10):341-51.

(99) Stenlund H, Madsen R, Vivi A, Calderisi M, Lundstedt T, Tassini M, et al. Monitoring kidney-transplant patients using metabolomics and dynamic modeling. Chemometr Intell Lab Syst 2009;98(1):45-50.

(100) Stenlund H, Gorzsís A, Persson P, Sundberg B, Trygg J. Orthogonal projections to latent structures discriminant analysis modeling on in situ FT-IR spectral imaging of liver tissue for identifying sources of variability. Anal Chem 2008;80(18):6898-906.

(101) Sadeghi-Bazargani H, Mehdizadeh R, Fathi K, Savadi-Oskouei D, Farhoudi M. application of PLS and OPLS modeling techniques to study the possible predictors of stroke prognosis after six months. Calcutta, India 2009.

(102) Pohjanen E, Thysell E, Jonsson P, Eklund C, Silfver A, Carlsson IB, et al. A multivariate screening strategy for investigating metabolic effects of strenuous physical exercise in human serum. J Proteome Res 2007 Jun;6(6):2113-20.

(103) Bergemalm D, Forsberg K, Jonsson PA, Graffmo KS, Brannstrom T, Andersen PM, et al. Changes in the spinal cord proteome of an amyotrophic lateral sclerosis murine model determined by differential in-gel electrophoresis. Mol Cell Proteomics 2009 Jun;8(6):1306-17.

(104) Eriksson L, Trygg J, Wold S. CV-ANOVA for significance testing of PLS and OPLS-« models. J Chemometr 2008;22(11-12):594-600.

(105) Colli R, Russo F, Bianchi G. [Minor burns in childhood. Clinical cases]. Minerva Pediatr 1991 May;43(5):371-5.

(106) Ehrlich AR, Kathpalia S, Boyarsky Y, Schechter A, Bijur P. Elderly patients discharged home from the emergency department with minor burns. Burns 2005 Sep;31(6):717-20.

(107) Glasheen WP, Attinger EO, Anne A, Boyd DR, Ruffin W, Haynes B, et al. Epidemiology of minor burn injuries. Burns Incl Therm Inj 1982 Jul;8(6):423-32.

(108) Han TH, Kim JH, Yang MS, Han KW, Han SH, Jung JA, et al. A retrospective analysis of 19,157 burns patients: 18-year experience from Hallym Burn Center in Seoul, Korea. Burns 2005 Jun;31(4):465-70.

(109) Rea S, Wood F. Minor burn injuries in adults presenting to the regional burns unit in Western Australia: a prospective descriptive study. Burns 2005 Dec;31(8):1035-40.

(110) Trilling JS. Minor burns: a review. Fam Pract 1987 Sep;4(3):235-9.

(111) Bristol N. US region to model health service on Iranian system. Lancet 2010 Feb 20;375(9715):625.
(112) Marandi A. Integrating medical education and health services: the Iranian experience. Med Educ 1996 Jan;30(1):4-8.
(113) Asadi-Lari M, Sayyari A, Akbari M, Gray D. *Public health improvement in Iran--lessons from the last 20 years*. Public Health 2004;118(6):395-402.
(114) Mohammadi R, Sadeghi-Bazargani H, Gooya MM. Burns to young working females in work and home settings. Box in. World report on child injury prevention. Peden,M; Oyegbite,K; Ozanne-smith,J; Hyder,AA; Branche,C; Fazlur Rahman,AKM; Rivara,F; Bartolomeos,K ed. World health organization; 2008. p. 82.
(115) Davies JW. The problems of burns in India. Burns 1990;Suppl 1:S1-24.
(116) Forjuoh SN, Guyer B, Smith GS. Childhood burns in Ghana: epidemiological characteristics and home-based treatment. Burns 1995 Feb;21(1):24-8.
(117) Peck MD, Kruger GE, van der Merwe AE, Godakumbura W, Ahuja RB. Burns and fires from non-electric domestic appliances in low and middle income countries Part I. The scope of the problem. Burns 2008 May;34(3):303-11.
(118) Peck MD, Kruger GE, van der Merwe AE, Godakumbura W, Oen IM, Swart D, et al. Burns and injuries from non-electric-appliance fires in low- and middle-income countries Part II. A strategy for intervention using the Haddon Matrix. Burns 2008 May;34(3):312-9.
(119) Othman N, Kendrick D. Epidemiology of burn injuries in the East Mediterranean Region: a systematic review. BMC Public Health 2010;10:83.
(120) Simons MA, Ziviani J, Copley J. Predicting functional outcome for children on admission after burn injury: do parents hold the key? J Burn Care Res 2010 Sep;31(5):750-65.
(121) Oster C, Kildal M, Ekselius L. Return to work after burn injury: burn-injured individuals' perception of barriers and facilitators. J Burn Care Res 2010 Jul;31(4):540-50.
(122) Van NA, Menckel E, Laflamme L. Barriers and enablers to the use of measures to prevent pediatric scalding in Cape Town, South Africa. Public Health Nurs 2010 May;27(3):203-20.
(123) Tsai JH. Chinese immigrant restaurant workers' injury and illness experiences. Arch Environ Occup Health 2009;64(2):107-14.
(124) Phillips C, Fussell A, Rumsey N. Considerations for psychosocial support following burn injury--a family perspective. Burns 2007 Dec;33(8):986-94.
(125) Munro SA, Van NA, Seedat M. Childhood unintentional injuries: the perceived impact of the environment, lack of supervision and child characteristics. Child Care Health Dev 2006 May;32(3):269-79.
(126) Thompson R, Boyle D, Teel C, Wambach K, Cramer A. A qualitative analysis of family member needs and concerns in the population of patients with burns. J Burn Care Rehabil 1999 Nov;20(6):487-96.
(127) Nagy S. Strategies used by burns nurses to cope with the infliction of pain on patients. J Adv Nurs 1999 Jun;29(6):1427-33.
(128) Anderson FD, Maloney JP, Redland AR. Study of hope in patients with critical burn injuries. J Burn Care Rehabil 1993 Mar;14(2 Pt 1):207-14.

(129) Forjouh S, Geilen A. Burns. In: Peden M, Oyegbite K, Ozanne-smith J, Hyder A, Branche C, Fazlur Rahman A, et al., editors. World report on child injury prevention. World health organization; 2008. p. 79-100.
(130) Walender G, Svanstrom L, Ekman R. Safety promotion-an introduction. Stockholm: Karolinska institute; 2004.
(131) Masud T, Morris RO. Epidemiology of falls. Age Ageing 2001 Nov;30 Suppl 4:3-7.

12 APPENDICES

Appendix 1:

Related publications of the student not included in the thesis

1- Mohammadi R, Sadeghi-Bazargani H, Gooya MM. Burns to young working females in work and home settings. Box in. World Report on Child Injury Prevention. Peden,M; Oyegbite,K; Ozanne-smith,J; Hyder,AA; Branche,C; Fazlur Rahman,AKM; Rivara,F; Bartolomeos,K ed. World Health Organization; 2008. p. 82.
2- Arshi S, Sadeghi-Bazargani H, Mohammadi R, Mohammad Zadeh MS, Rouhi A, Barak M. Burns comprising 3/4th of home injuries in pre-school children of the rural areas of Ardabil Province, Iran. J Med Sci 2007;7(2):248-51.
3- Arshi S, Sadeghi-Bazargani H, Mohammadi R, Ekman R, Hudson D, Djafarzadeh H, et al. Prevention oriented epidemiologic study of accidental burns in rural areas of Ardabil, Iran. Burns 2006 May;32(3):366-71.
4- Sadeghi-Bazargani, H., et al. "The risks of using samovars as the main tea-preparing facility in some Eastern countries." Burns 34.8 (2008): 1149-52.

Appendix 2

The mathematical algorithm developed by J.Trygg for OPLS in 2006 as referred to in the text:**(77).**

1. $w^T = y^T X \ / \ (y^T y)$.
2. $w = w \ / \ \|w\|$.
3. $t = Xw \ / \ (w^T w)$.
4. $c^T = t^T y \ / \ (t^T t)$.
5. $u = yc \ / \ (c^T c)$.
6. $p^T = t^T X \ / \ (t^T t)$.
7. $w_{ortho} = p - [w^T p \ / \ (w^T w)]w$.
8. $w_{ortho} = w_{ortho} \ / \ \|w_{ortho}\|$.
9. $t_{ortho} = Xw_{ortho}/(w^T_{ortho} w_{ortho})$.
10. $p^T_{ortho} = t^T_{ortho} X/(t^T_{ortho} t_{ortho})$.
11. $E_{O-PLS} = X - t_{ortho} p^T_{ortho}$.
12. Save found parameters $T_{ortho} = [T_{ortho} \quad t_{ortho}]$, $P_{ortho} = [P_{ortho} \quad p_{ortho}]$, $W_{ortho} = [W_{ortho} \quad w_{ortho}]$.
13. $X_{ortho} = T_{ortho} P^T_{ortho}$.
14. $X_{ortho} = T_{pca\_ortho} P^T_{pca\_ortho} + E_{pca\_ortho}$.