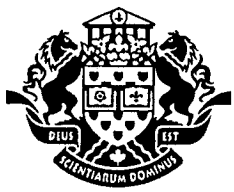
NOTE TO USERS



Université d'Ottawa · University of Ottawa

Université d'Ottawa · University of Ottawa

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES

FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

Janet BERKMAN

AUTEUR DE LA THÈSE - AUTHOR OF THESIS

M. Sc. (Mathematics spec. Biostatistics)

GRADE - DEGREE

Department of Mathematics and statistics

FACULTÉ, ÉCOLE, DÉPARTEMENT - FACULTY, SCHOOL, DEPARTMENT

TITRE DE LA THÈSE - TITLE OF THE THESIS

Predicting patient knowledge of cardiac risk factors : a comparison of two approaches

A. Dabrowski

DIRECTEUR DE LA THÈSE - THESIS SUPERVISOR

CO-DIRECTEUR DE LA THÈSE - THESIS CO-SUPERVISOR

EXAMINATEURS DE LA THÈSE - THESIS EXAMINERS

Y. Chen

A. Mayer

M. Salibian-Barrera

J.-M. De Koninck, Ph.D.

LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES

DEAN OF THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

PREDICTING PATIENT KNOWLEDGE OF CARDIAC RISK FACTORS: A COMPARISON OF TWO APPROACHES

JANET BERKMAN

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
University of Ottawa
In partial fulfillment of the requirements for the
M.Sc. degree in the

Ottawa-Carleton Institute for Graduate Studies and Research in Mathematics and Statistics

Your file Votre référence
ISBN: 0-494-01416-4
Our file Notre référence
ISBN: 0-494-01416-4

Abstract

The University of Ottawa Heart Institute conducted a survey of patients to understand the level of knowledge of cardiac risk factors and to identify any subgroups of patients that could benefit from specially designed educational programs. This thesis compares two approaches to the analysis of this multidimensional dataset.

Both techniques looked at the 10 modifiable risk factors and a number of predictor variables (age, gender, education, and smoking status). Logistic regression was hampered by low sample size, sparse data, and the high probability responses of many of the binary knowledge variables. Only one risk factor was successfully explained by any of the predictor variables. Correspondence analysis demonstrated that those who are unaware of smoking as a risk factor are not current smokers; knowledge of low fibre diet is related to education but not to gender; females are more aware of high salt diet and stress as risk factors, and are more likely to have never smoked; smokers tend to have lower education and be unaware of the risk of a low fibre diet.

Acknowledgements

In divesting itself of thousands of employees in the past few years, Nortel Networks provided the impetus (and means) for many of us to start something new. After a 17 year diversion into telecommunications, I was able to return to a field that first sparked my interest in a course taught by Professor David Andrews at the University of Toronto, back in 1983, "Special Topics in Biostatistics".

I know that my approach to data analysis has been forever modified by my studies in epidemiology. I now appreciate the inherent beauty in a 2x2 table. I discovered that statistical significance is not necessarily significant, and thought a lot about causality. I will never again exclude an "outlier" without thinking twice about it. Thanks to Professors Spasoff, Wells, and Chen in the Faculty of Medicine, Department of Epidemiology and Community Medicine, University of Ottawa.

I am grateful to many people who have made this thesis possible. Dr. Kathryn Momtahan was a colleague of mine in human factors at Nortel Networks. As she returned to a research position in Nursing at the University of Ottawa Heart Institute, I was looking for some data to examine for this thesis. I was made to feel part of her team at the UOHI, and I appreciate her willingness to work remotely with me as I moved out of Ottawa. I hope that we can continue to collaborate in the analysis of this very rich dataset.

My supervisor Dr. Andre Dabrowski has supported me unfailingly through this degree. He, too, was willing to work with me remotely on this thesis. He put up with delays in the original schedule as I diverted energies away from the main thrust of the thesis to co-author and publish the journal article that appears here.

Finally, to my dear husband Zouheir Mansourati, who knew just how much he could ask me about my progress without being annoying.. He has always believed in me and supported me, and for that I am forever grateful. To my children Alexander and Michael, who make me laugh and keep me grounded.

Table of Contents

1. Introduction

Cardiovascular diseases are one of the leading causes of death in Canada. Much research effort is directed to new and/or improved treatments for these diseases as well as to increasing knowledge of risk factors and methods of prevention. This thesis examines patient education, a relatively sparsely studied domain.

The University of Ottawa Heart Institute (UOHI) provides cardiovascular care services to residents of Eastern Ontario. These services include a comprehensive range of diagnostic procedures as well as medical and surgical treatment. The Institute also offers rehabilitation and educational programs to support primary and secondary prevention. Many of these services are coordinated with other public health service providers such as the regional health authorities. The Institute also houses a 90,000 square foot Research Centre.

The Clinical Services Department (CSD) at the UOHI provides educational programs to current and former patients in the form of lectures, courses, and written materials. To better design and target these services, the CSD initiated a study in 2002 to understand the level of knowledge of cardiac risk factors in their patient population (Research Goal 1). In addition, they were interested in identifying whether there are sub-groups of patients that could benefit from specifically designed educational initiatives (Research Goal 2). After reviewing the research goals and having the opportunity to join the research team at the early stages of the study, this project seemed an ideal source of data for this thesis. It provided a very rich, self-contained dataset that would provide the basis for discussion of statistical issues and alternative analytical approaches. This led to the development of the Thesis Goal: to compare two approaches to the analysis of this multi-dimensional dataset: logistic regression modeling and correspondence analysis.

There is scant literature at present regarding patient knowledge of risk factors. Chapter 2 presents further information on the background of this study and summarizes the few papers in which this subject has been studied previously.

Chapter 3 describes the study design. This was unique in that it involved a threepronged approach: both patients and their significant others were interviewed separately, and a chart audit was performed to capture information about health status, including blood chemistry and medication profiles.

The first part of the analysis addresses Research Goal 1. It is descriptive in nature and the results are summarized in a paper that was published in the Journal ofCardiovascular Nursing[1], reprinted here in Appendix 4. Additional descriptive analysis is also contained in Chapter 4 of this thesis.

The bulk of the thesis addresses Research Goal 2 (and the Thesis Goal): Which, if any, subgroups of the UOHI patient population have particular educational requirements vis-à-vis cardiac risk factor education? We decided to take two different analytical approaches that we felt wouldbe complementary. With a relatively small sample size of 71 patients, multidimensional data, and the exploratory nature of this research, we were interested in attempting both a quantitativemodeling approach (logistic regression) and a qualitative or data-visualization approach (correspondence analysis).

Chapter 5 introduces and describes these approaches. In the case of logistic regression, we modeled patient knowledge of individual risk factors using various demographic variables as predictors. The advantage to this approach is that it produces a quantifiable relationship between the outcome and predictors. The disadvantage of standard logistic regression is that we can only examine one risk factor at a time. We lose the richness of the data in whicheach patient has (or

lacks) knowledge of a suite of risk factors and it seems probable that a patient's knowledge of one risk factor is not independent of their knowledge of another risk factor.

The second approach examines groups of risk factors and demographic variables together, a decided advantage over the first approach, but with the disadvantage of no quantitative descriptors of relationships. Rather, correspondence analysis produces graphs that represent the "closeness" of the variables (both knowledge variables and demographic variables) in multidimensional space. There is no longer a distinction mathematically between outcome and predictor variables. Sections 5.4 and 5.5 compare these two approaches and discuss implications for future experimental design (Thesis Goal).

Chapter 6 assesses how well we addressed the research and thesis goals and suggests areas for future research.

2. Background

Cardiovascular diseases[2] accounted for 37% of deaths in Canada in 1996.[3] Age-adjusted mortality is 226 deaths per 100,000 population, larger than any other disease group. Hospital separations for circulatory diseases are second only to pregnancy-related conditions. However, in Canada, the mortality rate has been declining since the mid-1960s, partly due to reductions in the level of risk factors, including smoking, consumption of dietary fat, management of hypertension, as well as improvements in the medical management of these diseases. A study published in 1999 suggests that 49% of Canadian adults between the ages of 55 and 74 have at least three of the major modifiable risk factors (high blood pressure, smoking, high cholesterol, inactivity, and obesity)[4].

Continued reductions in morbidity and mortality rates can be effected by on-going efforts in education regarding risk factors targeted at patients and their health care providers. This thesis will describe a study conducted in 2002 at the University of Ottawa Heart Institute. The study was designed to look at knowledge of cardiac risk factors (CRFs), from both general and personal perspectives, for patients admitted to the Heart Institute. It examined the knowledge levels of both the patient and the patient's significant other. The objective of this thesis is to examine the relationships between a patient's general knowledge of CRFs and their personal characteristics, including demographic variables (e.g., gender, age, education, diagnostic category, physical characteristics, and language) and sources of risk factor information. In the process, we will contrast two statistical methods of presenting the data, and highlight their separate advantages.

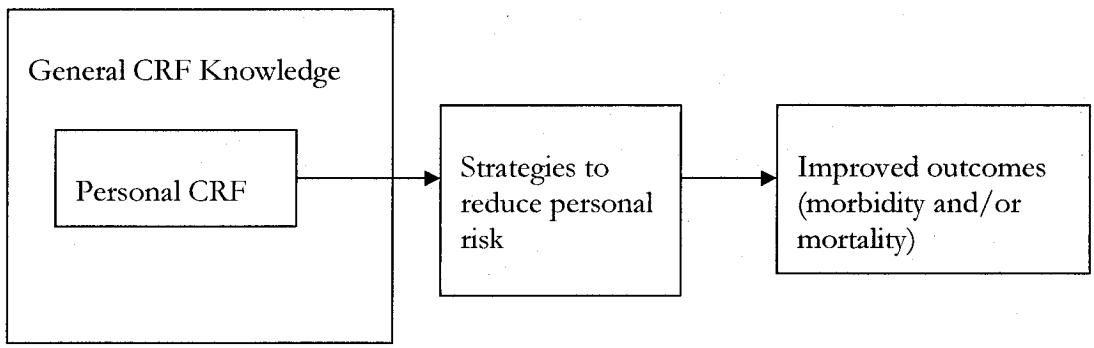
2.1. Theoretical framework

A Common-Sense Model (CSM) of illness representations was first proposed by Leventhal et al in 1980[5]. It describes how individuals organize information about illness, make decisions about managing their illness, and the resulting outcomes, both physical and emotional. In the CSM, illness representations are guided by three basic sources of information: (1) lay information or general knowledge; (2) information gathered from authoritative sources such as a doctor, parent or other significant individual; and (3) the patient's own experience with the illness (including symptoms, efficacy of treatment, etc),

In this study we will be focusing on the first two information sources but have somewhat reframed the classification. We ask patients about both their general knowledge of cardiac risk factors as well as their knowledge of their own personal risk factors.

In Figure 1 , we suggest a heuristic model of risk factor knowledge and management.

Figure 1 - A heuristic model describing how general and personal information on CRFs are related, and could influence morbidity



In this model, General CRF Knowledge can come from many sources: medical professionals (personal physician, nurse, dietician, health educator), other people (friends, family), or media (books, television, internet, radio, pamphlets). General knowledge of CRFs can be applied to oneself, family members, or friends, and can also be used in policy making (for example, smoking policies, community activity programs) or in other ways (for example, design of restaurant menus or workplace health programs).

Personal CRF knowledge can come from either one's own awareness that general CRFs apply to oneself, or from medical professionals acting to assess health status or obtain a diagnosis. General CRF knowledge may lead to personal CRF knowledge, and certainly personal CRF knowledge leads to general knowledge (e.g., if I learn that I have low HDL cholesterol and that this is a CRF, I can assume that other people with low HDL are also at increased risk for heart disease.)

So we suggest that general knowledge of CRFs is a necessary condition for individuals to work towards improving or maintaining their health status. Even in the absence of existing disease, primary prevention requires a forward looking approach to lifestyle changes, as many of the modifiable risk factors are developed over a long period of time. Diet and physical activity have long-term impacts on risk status and general knowledge gained early in life is more likely to have the desired impact on lifestyle and behaviour.

2.2. Risk factors for heart disease

Risk factors for heart disease can be grouped into two categories: modifiable and non-modifiable (see Table 1). Reduction of exposure to these risks can substantially reduce morbidity and mortality. Extensive work has been done to quantify the relationship between risk factors and

coronary heart disease (CHD) risk, most notably by the Framingham investigators[6]. This thesis will not look at the relationship between risk factors and heart disease, but between patient demographics and risk factor knowledge.

Table 1 - Risk Factors for Heart Disease

Modifiable	Non-Modifiable
Cholesterol (High LDL, Low HDL)	Age
Hypertension	Family history
Smoking (Current or Former)	Gender
High blood sugar/Diabetes	Passive smoking (exposure to second hand smoke)
Obesity	
Inactivity	
High salt diet	
Low fibre diet	
High fat diet	
Stress	

2.3. Studies of risk factor knowledge

A number of studies have examined the level of general and/or personal risk factor knowledge. Some of these studies ask patients to name cardiac risk factors with no assistance from a list. We will refer to these as "recall tasks". Another approach would be to provide patients with a list of possible risk factors and ask them to indicate whether each item is a cardiac risk factor. We will

refer to these as "recognition tasks". The study described in this thesis contained both recall and recognition tasks.

2.3.1. Recall studies

In a 1978 study by Shekelle and Liu[7], more than half of the 617 adult subjects could not name any of the three main causes of heart attacks (high blood pressure, cigarette smoking, high cholesterol or fat in the diet).

The most frequently mentioned cause was stress (36%) followed by overweight (30%), cigarette smoking (28%), and hypertension (21%). High cholesterol or fat in the diet or blood was named by 13%. Shekelle and Liu noted that the prevalence of these responses was positively related to education (although there is no level of significance provided, suggesting that this finding was of a qualitative nature). For example, cigarette smoking was named by 24% of those without high school education, 28% of high school graduates, and 32% of college grads.

In 1988, Folsom et al[8] reported that 70% of the Chicago-area adults in their study could correctly identify at least one of the three main modifiable risk factors. This paper also noted a relationship between educational level and beliefs about cardiovascular disease. While hypertension and over exertion were more commonly mentioned by less educated persons, the remaining risk factors were mentioned with increasing frequency as education levels increased in both men and women. Women in this study were more likely to mention overweight, hypertension, diet, and heredity and less likely to mention smoking and over-exertion. All of these findings were stated as "$p<.05$" although the statistical test used was not specified.

Zerwic et al (1997)[9] surveyed 105 patients hospitalized for myocardial infraction or for coronary angiography with the diagnosis of coronary artery disease, asking open-ended questions regarding

cardiac risk factors. They found that 79% of the patients named at least one of three modifiable risk factors (smoking, hypertension, elevated cholesterol) but only 7% identified all three. Twelve percent could not name any risk factors. The most frequently cited causes were diet (67%) and smoking (50%).

Finally, a Canadian study published in 1999 by Kirkland et al[10] analysed a subset of the data from the Canadian Heart Health Survey for 4976 subjects aged 55 to 74 not living in institutions. In order of frequency mentioned, recall of risk factors were as follows: stress or worry (44%), smoking (41%), poor diet (37%), excess weight (30%), lack of exercise (28%), heredity (24%), high blood cholesterol (23%), excess fats (21%), overwork (21%), hypertension (16%), high cholesterol foods (10%) and excess salt (6%). More women than men recalled hypertension and heredity, and there was lower overall awareness of risk factors in the upper half of the age range (65-74). Most differences were found to be statistically significant, but this was due to the large sample size so significance levels were not reported and discussion of differences was limited to those deemed clinically important (for example, differences between men and women and between age groups as mentioned above.)

2.3.2. Recognition studies

In 1985, the National Health Interview Survey[11] in the US included questions on risk factors for heart disease. In this recognition task, 94% of respondents stated that overweight is a risk factor. Other factors included in the task were high blood pressure (92%), smoking (91%), high cholesterol (87%), family history (83%), high animal fat in the diet (80%), and diabetes (60%).

We were unable to identify any other studies that took a recognition approach to studying CRF knowledge.

3. Study design for the UOHI survey

3.1. Population

The population for this study consisted of all patients who were hospitalized at the University of Ottawa Heart Institute (UOHI) on a Wednesday in June 2002. The UOHI is a teaching hospital that provides bilingual French/English cardiac services to a region consisting of over 1.8 million residents. Patients in all units except the operating rooms, the cardiac recovery room, the cardiovascular ICU, and the cardiac catheterization labs were asked to participate in the study along with their significant other (SO). The exclusion criteria included patients who were not literate in English or French and did not have someone to translate for them, patients who were confused, deemed unable to participate due to their condition (for instance, were mechanically ventilated or deemed too ill by the interviewer to participate), and those who were assessedto be too close in time to a procedure or operation to participate. On this basis, 26 patients were excluded from the study. Eight patients who may have been willing to participate were missed due to timing issues (for example, they were discharged suddenly). After accounting for excluded and missed patients, this left 87 patients who were able to participate. Thirteen declined to participate and three patients decided not to participate after they had signed the consents. The patient response rate was 75%, that is, the number of surveys completed as a fraction of the number of eligible patients (71/(121-26)).

Conducting all the patient surveys in a single day has both advantages and disadvantages. Single day collection meant that a team of nurses could be made available at one time and trained just prior to the interviews. It also meant that hospital staff was made aware of the study in a timely manner, and that patients could be informed the evening prior to the study day (or first thing in the morning). On the other hand, single-day collection meant that any seasonality in admission

patterns were missed, possibly introducing some sampling bias. Considering the goals of the thesis, the advantages of single-day collection outweighed other sampling designs.

The SO was defined as the person that the patient thought would be most involved with their care and recovery once they left the hospital. The SO was usually a spouse but could be another family member or a close friend. The decision was made to survey the SOs as well as the patients for two reasons. First, family and friends may be instrumental in helping patients recover and modify their lifestyles once they return home[12]. Second, we wanted to determine the congruence among the answers given by the patients and their SOs regarding the patient's cardiac risk factors. Not all patients were able to identify an SO and some SOs declined to participate, resulting in a sample size of 53 for the SO interviews.

3.2. Procedure

There were three parts to the study: the patient interviews, the SO interviews, and the chart audits. All of the interviewers and chart auditors were nurses employed at the UOHI.Ethics approval was obtained from the Heart Institute Human Research Ethics Committee (see Appendix 1).

3.2.1. Interviews

Twelve interviewers interviewed the 71 patients and the majority of the 53 SOs during the 12 hour study period. Interviewers were assigned to patients based on language and location in the UOHI. Patients were interviewed in as private a manneras possible for reasons of patient confidentiality as well as to maintain independence of responses between patients. Some of the SOs were not available to be interviewed in hospital and were later contacted by telephone. The patient interviews and the SO interviews were conducted separately tokeep the responses individual and to maintain the integrity of the recall task. The interviewers participated in two three-hour training sessions prior to the interview day. Inter-rater reliability testing was done to achieve 85% agreement on the information gathered from mock interviews*.

Patients and their SOs were asked open-ended questions first, to determine their level of unaided awareness of cardiac risk factors. Following these open-ended questions, patients and their SOs were presented with what was described as a list of possible cardiac risk factors. Sixteen were in

---

* Inter-rater reliability was achieved by training the interviewers as a group, and then performing two mock interviews. A nurse played the role of two different patients (surgical and medical) and trainees completed survey forms while observing the principal investigator conduct the interviews. Completed survey forms were reviewed as a group to ensure a high level of reliability in recording the interview data.

fact risk factors, while 9 of these were not cardiac risk factors but sham risk factors[‡]. They were asked to identify whether each item in the list was a cardiac risk factor in general and if the answer was "yes", they were asked if this was a risk factor that they personally possessed. The patient and significant other interview instruments are contained in Appendices 1 and 2 respectively along with the patient information sheet and consent form.

3.2.2. Chart Audits

Chart audits were performed over a two-month period post discharge after the charts had been received in medical records. Ten nurses audited the completed charts of the 71 patients who were interviewed. Four nurses who had also participated in the interviewing of patients and their SOs audited 82% of the charts. To ensure that the charts were reviewed consistently, these four auditors did the first day of chart auditing in a group. Since the information gathered for the chart audits was much less ambiguous than the information gathered by the interviews, no formal inter-rater reliability assessment was done. The chart audit data collection form is contained in Appendix 3.

[‡] We acknowledge that, technically, we have only lack of evidence that these so called "sham" risk factors were in fact NOT risk factors for heart disease. We refer the reader to the criteria suggested by Hill for attempting to distinguish causal from noncausal associations. See Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295-300.

4. Descriptive analysis of UOHI data

4.1. Recall and recognition of general risk factors

Patients were asked to name some risk factors for heart disease. The 71 patients interviewed mentioned relatively few factors in this simple recall task (Table 2).

Table 2 - Unprompted Mentions of Cardiac Risk Factors (Percents do not add to 100 as multiple mentions were encouraged. N=71)

Risk Factors	Unprompted Mentions	Percent
Smoking	50	70
Lack of Physical Activity	21	30
Family History of Heart Disease	21	30
Stress	19	27
High Fat Diet	16	23
Overweight	15	21
High Blood Cholesterol	12	17
High Blood Sugar	10	14
High Blood Pressure	5	7
High Salt Diet	3	4
Low Fibre Diet	1	1

However, when presented with a list of possible risk factors and asked to indicate whether or not each factor was in fact a CRF (recognition task), patients did much better (Table 3). For example, while 70% of patients mentioned smoking as a risk factor when asked to recall risk factors (unprompted), 97% recognized correctly identified smoking when it was mentioned. Similarly, only 17% could recall that high cholesterol is a risk factor, but 96% of patients recognized it as a risk factor when prompted.

Table 3 - Risk Factor Knowledge When Prompted

Risk Factors	Q: Is this a RF for heart disease?			
	Yes	No	Don't Know	Total
Modifiable				
High blood pressure	69	0	2	71
High blood cholesterol	68	0	3	71
Overweight	68	2	1	71
Smoking	68	2	1	71
High fat diet	66	2	2	70
Stress	65	4	2	71
High salt diet	61	6	4	71
Lack of physical activity	61	7	3	71
High blood sugar	59	5	7	71
Former smoker	57	9	5	71
Low fibre diet	35	20	16	71
Non-modifiable				
Family history of heart disease	68	2	1	71
Previous heart problem	66	0	5	71
Passive smoking	63	5	3	71
Male > age 45, female > 55	57	11	3	71
Sham risk factors				
History of rheumatic fever	38	15	17	70
Male > 35, female > 55	36	27	8	71
Gout	17	26	28	71
Having had previous surgery	16	40	15	71
History of cancer	14	34	23	71
Having had a general anaesthetic	7	41	23	71
Hayfever	6	35	30	71
History of broken leg	5	61	5	71
Having had chicken pox	4	38	29	71

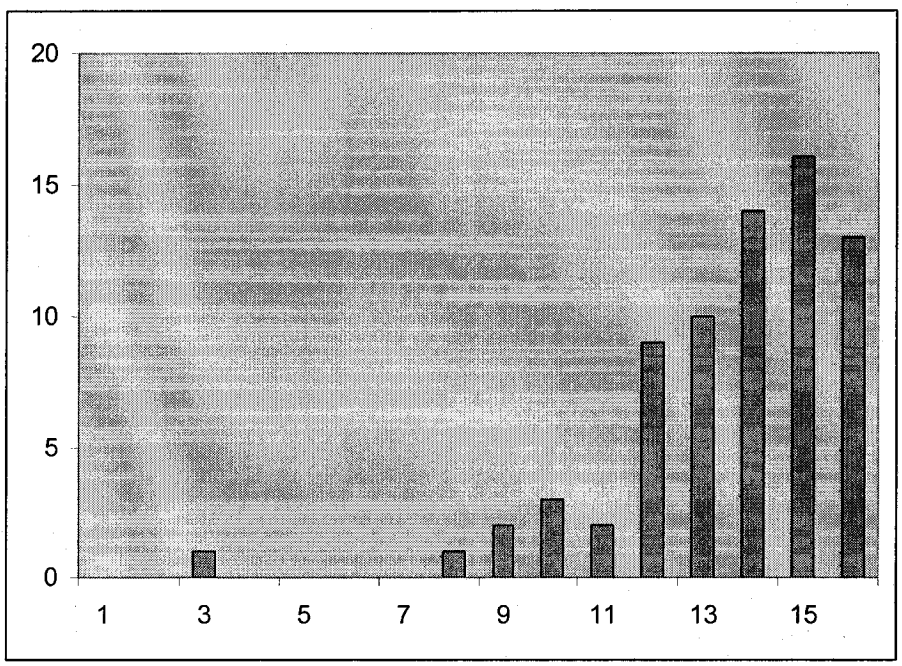
4.2. Risk Factor Knowledge Score

By summing the number of correct ("yes") answers to the 16 modifiable and non-modifiable risk factor questions, we created an index, the RFI, to proxy the patients' general knowledge of CRFs. The overall mean general knowledge score was 13.6 out of a possible 16 (s.d.=2.3; min,max::3,16).[‡]

We then wished to assess whether there were differences in knowledge score across subgroups. We began by examining histograms of the RFI data. Figure 2 contains a histogram of the RFI data and we note that the scores are skewed to the left, with the majority of patients scoring at least 12 out of 16. We also used a T-test for equivalence of means however the assumption of normality of the underlying distribution is questionable. Observations are by definition equal to or less than 16, with most observations occurring at the upper end of the range (0,16).

[‡] An alternative measure of risk factor knowledge is RFI-S = [#correct for true risk factors] – [# incorrect for sham risk factors]. For this variable, the mean score was 11.5 (2.3;3,15) out of a possible maximum score of 16. We choose to concentrate on the former or uncorrected RFI, that is, on positive knowledge.

Figure 2 - Histogram of Risk Factor Index (RFI)



We next examined histograms for various subgroups of the data. Figure 3 compares scores for males and females. We note that the data for males appears to have a wider spread with more males scoring under 12. Males scored on average 13.2 and females scored 14.2. The T-test for equivalence of means yielded $p=0.08$. While this difference is not statistically significant at the usual $\alpha=0.05$ level, this is a suggestive result.

Figure 3 - Histogram of Risk Factor Index by gender.
Note the generally lower scores for males.

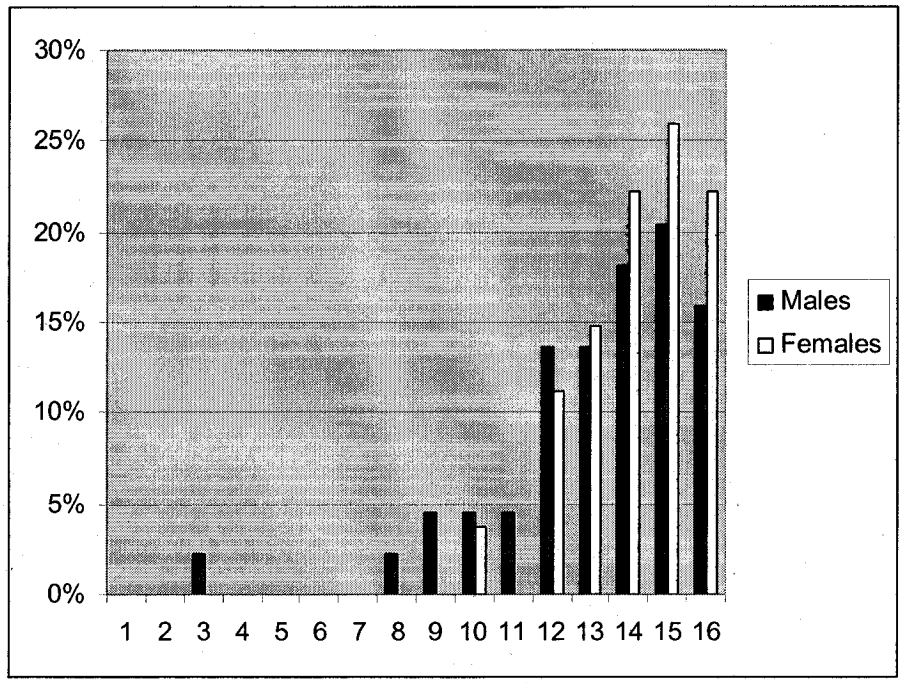
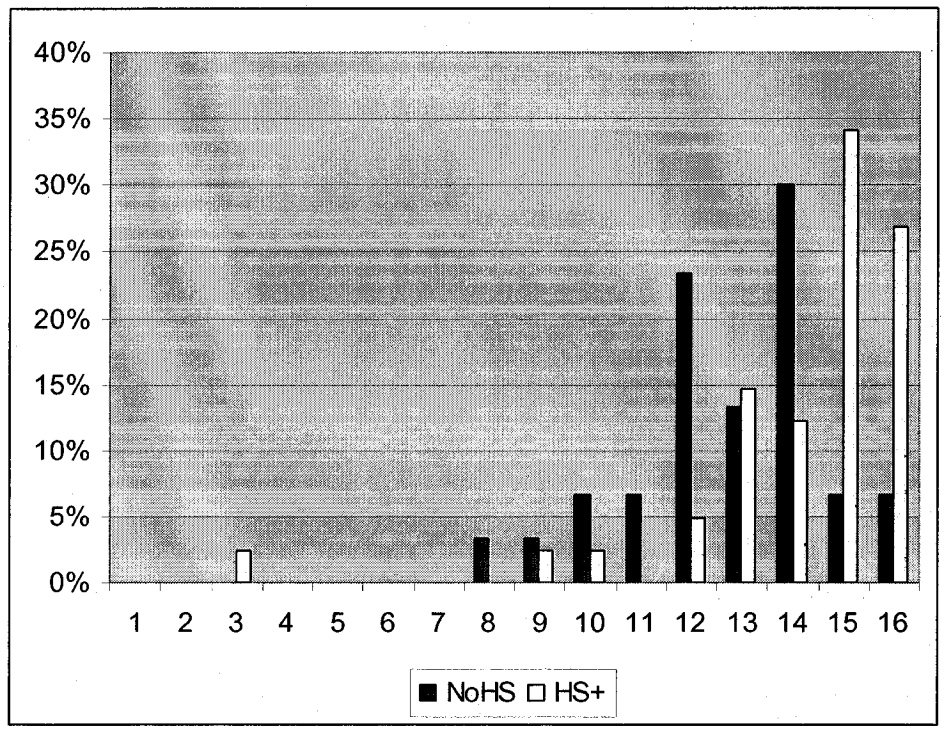


Figure 4 compares scores for patients with and without a high school diploma. We note that the distribution of responses for those with higher education is more skewed, suggesting that education plays a role in patients' risk factor knowledge. Those with high school education or greater scored significantly higher than those who did not complete high school. (14.2 vs 12.8; p=0.012.) Examining college and university-educated individuals separately from those with high school did not yield significant differences.

Figure 4- Histogram of Risk Factor Index by education. We note that patients with high school (HS) or greater education have generally higher scores.



There was no difference in mean score between groups of patients with and without previous heart problems (13.5 vs. 13.7).

Mean knowledge scores by age group were as follows: 30s and 40s: 12.9; 50s & 60s: 14.0; 70s & 80s: 13.2. A scatter plot of the ungrouped data is contained inFigure 5. We note one very low RFI score for one of the more elderly patients.

Figure 5 – Scatter Plot of RFI by Age

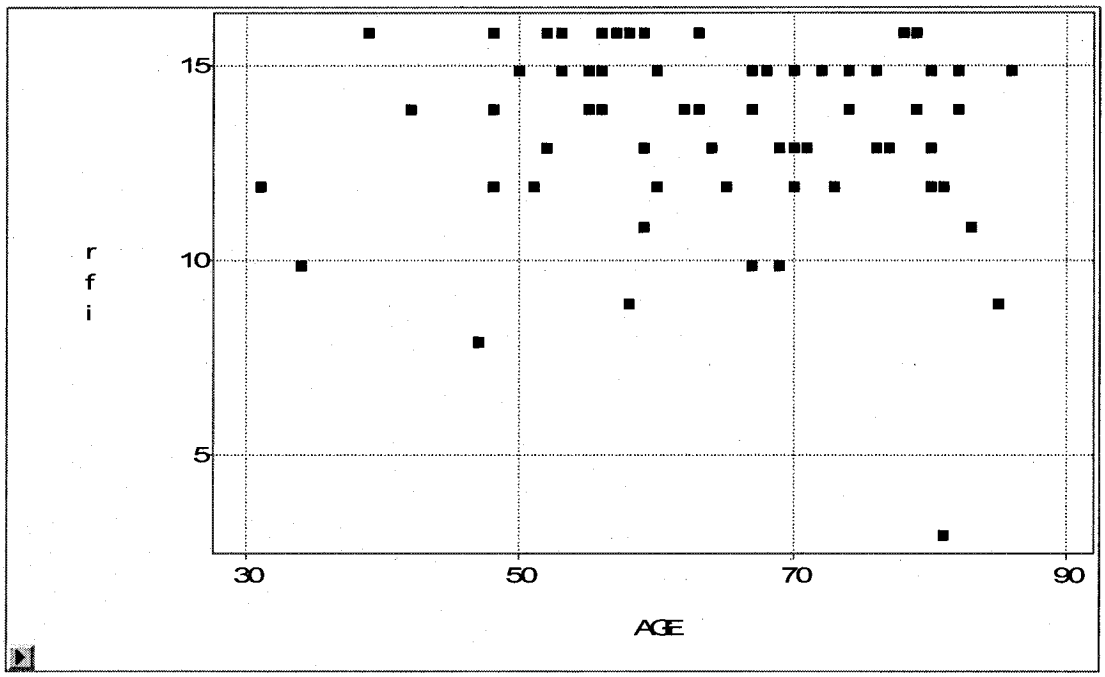
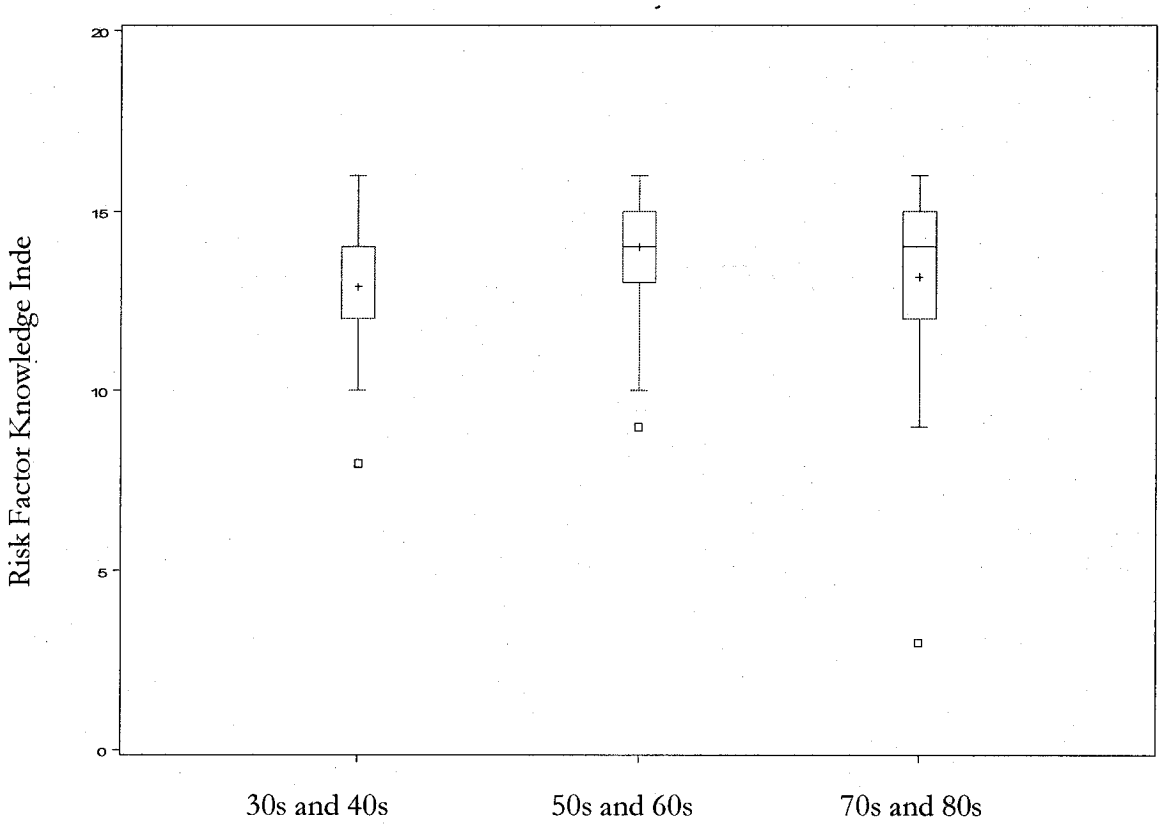


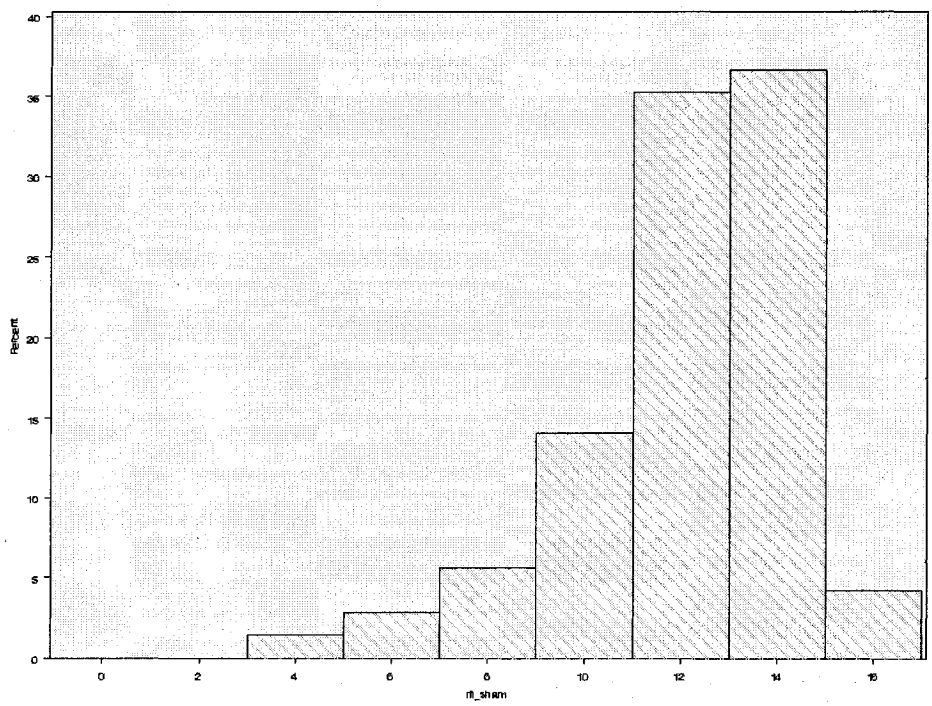
Figure 6 contains box plots of the RFI data by age group (30-49, 50-69, and 70-89). In this and subsequent box plots, the upper and lower limits of the box correspond to the first and third quartiles of the data (the inter-quartile range (IQR)). The line across the middle of the box is the median, and the cross in the box is the mean. The upper and lower lines correspond to the maximum observation within 1.5 IQRs of the 25th or 75th percentiles. The square symbols outside of the lines represent observations outside of these maximum observations. This demonstrates that the oldest patients had the most spread in their knowledge scores and the lowest scores of the three age groups. As well, the middle age group has the least variability and highest mean score.

Figure 6 - Box Plot of Risk Factor Index (RFI) by age group



We also examined the index formed by subtracting incorrect responses to the sham factors from the correct responses to the true risk factors. This yielded a variable "RFI-sham" that had a more bell-shaped curve (see histogram in Figure 7) but subgroup analysis as above yielded similar results.

Figure 7 - Histogram of RFI-sham.



4.3. Comparison of results with the literature

In Table 4, we compare results from this study with those of Shekelle, Folsom and Zerwic and examine the number of patients that identified the three major risk factors (smoking, high blood pressure, and high cholesterol). All four studies asked the question as a recall task. In the recognition task, virtually all patients (70 out of 71) were able to identify at least one and fully 93% of patients identified all three major risk factors.

Table 4 - Percentage of patients identifying major risk factors in CRF recall and recognition studies

Study	Population	Type of Questioning	Identified at least one major RF	Identified all 3 major RF
Shekelle & Liu (1978)	Community	Recall	49%	1%
Folsom et al (1988)	Community	Recall	70%	3%
Zerwic et al (1997)	Hospital	Recall	79%	7%
UOHI (2002)	Hospital	Recall	76%	4.2%
NHIS (1985)		Recognition		>80%
UOHI (2002)	Hospital	Recognition	99%	93.0%

Examining the unaided awareness of the three major risk factors, we note that 82% of male patients and 67% of female patients were able to identify at least one of these three. Of those without high school education, 70% identified at least one, vs. 80.5% of those with high school or greater. Seventy-eight percent of middle-aged patients (30s and 40s) and 89% of those in their 50s and 60s were able to identify at least one of the three risk factors, whereas only 56% of the 25 patients over 70 years of age were able to. Univariate analyses may mask more complex relationships. We will need to look to multivariate techniques to untangle these relationships and better understand which demographic characteristics are associated with CRF knowledge.

5. Approaches to Multidimensional Analysis

5.1. Introduction

Our objective is to understand relationships between patient knowledge of cardiac risk factors and measurable characteristics of patients. Patient knowledge is captured in a set of binary variables corresponding to the set of modifiable risk factors identified in Table 1. Patient characteristics are described in both qualitative (education, gender, language) and continuous (age) type variables. We may also be interested in other patient characteristics captured in their charts: presence of hypertension (binary), BMI (continuous), smoking status (discrete), etc. In this thesis, we will compare two different approaches to understanding these relationships: modeling and data visualization.

Modeling is an attempt to describe the relationship between observed variables and predictors in a quantitative way, and we will use a type of log-linear model (logistic regression) for this dataset. Logistic regression[13,14] is a standard part of the toolbox for analysis of binary response data and in the following section, we model patient knowledge using patient characteristics as explanatory variables. Like all regression techniques, the advantages of logistic regression include the ability to assess the statistical significance of relationships between covariates and the independent variable. Confidence intervals for these parameter estimates can be calculated, and the model can be used to predict the response of future cases. One of the drawbacks of this approach is that it does not look at the patients' overall risk factor knowledge as captured by their scores on the entire set of risk factors. Rather, we can only model one response variable (i.e., knowledge of a single risk factor) at a time.

The data visualization technique we will use is called Correspondence Analysis (CA)[15]. This approach allows us to "see" the relationship between some or all of the 10 risk factors and patient

characteristics, and (hopefully) enable us to obtain a qualitative understanding of such relationships. The advantage of techniques such as this is their intuitive nature and ability to describe complex relationships in a single picture The drawback is that we do not obtain quantitative or numerical measures of these relationships.

5.2. Logistic Regression

The material from this section is based on Hosmer and Lemeshow[13] and Fienberg[16]. We provide only an overview of relevant topics here.

5.2.1. Theory

Logistic regression is a popular (and appropriate) method of modeling binary data. It assumes that the probability, π, of a "1" response is a function of a set of covariates or explanatory variables $\underline{x}$ and that this function takes the form

$$\pi(\underline{x}) = \frac{e^{x'\beta}}{1+e^{x'\beta}}$$

By taking the logit transformation of $\pi(x)$ we have

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right]$$

or

$$g(x) = x'\beta$$

which now looks like the familiar linear regression model. Fitting such a model to data is more complex than linear regression with Gaussian errors. We do not have "observed" values for π but rather see 1's and 0's and so least-squares fitting of parameters β is not appropriate. To obtain estimates of β we use the maximum likelihood principle. For logistic regression, maximizing the likelihood $L(\beta)$ is equivalent to solving

$$\sum [y_i - \pi(x_i)] = 0$$

and

$$\sum x_i [y_i - \pi(x_i)] = 0 .$$

These expressions are non-linear in β and require an iterative computational process to obtain a solution. Such processes are well-established in popular statistics software such as SAS[17].

The goal of model fitting using logistic regression is to balance the goodness-of-fit of the model to the data with parsimony in the number of parameters, and so we use the following approach. We compare models with and without a variable of interest, and determine whether there is significant improvement in the fit of the model by using the likelihood ratio test.

Specifically, let

$$G = -2 \ln \left[\frac{\text{likelihood of the model without the variable}}{\text{likelihood of the model including the variable}} \right]$$

It can be shown that G has an approximate chi-squared distribution under the hypothesis that the model without the variable is correct and that the total sample size is large[§] This, then, becomes the benchmark for inclusion of a variable of interest. When considering the addition of a new predictor, small changes in G mean that the new variable does not explain significantly more variability and so should the new predictor should not be included in the model.

5.2.2. A computational problem

One of the difficulties in using log-linear models is the problem of incomplete data, that is, where not every combination of the classification variables exists in the data set. This situation should be distinguished from two other similar problems: missing data and structural zeros.

Missing data is the situation in which, for example, one or more question responses are missing for certain individuals in an otherwise complete survey. There are a number of techniques for dealing with missing data, including imputation[18].

Structural zeros occur when certain combinations of classification variables cannot occur. For example, data on the incidence of surgical procedures by gender will include structural zeros for certain procedures (e.g., hysterectomies performed on males). Structural zeros restrict the models that are valid for a given dataset.

[§] Monte Carlo studies conducted by Larntz suggest that a sample size equal to four or five times the number of cells in the table (distinct combinations of levels of predictor variables) should be adequate for the use of the chi-square approximation. In this study, this requirement means that the number of cells should be less than or equal about 16. With 3 levels of smoking, 2 levels of gender, and 3 education groups, this is 18 distinct levels, meaning that our approximation may become somewhat strained when multiple predictors are included. See Larntz K. *Small sample comparisons of exact levels for chi-square goodness-of-fit statistics.* JASA. 1978;73:253-263.

In the situation at hand, we are talking about the case where, for reasons of small sample size (or equivalently, small probability responses), we simply have not observed a particular combination of explanatory variables. For example, in this dataset, there are no French speaking females who currently smoke. If we wish to understand the relationship between gender, smoking, language and RF knowledge, this data cannot tell us anything about French speaking smoking females, and so we cannot validly separate the effect of smoking from the effect of being a francophone female. In other words, the possible set of models to be distinguished is limited by the data.

5.2.3. Models for 10 major modifiable risk factors

We fit simple models for each of the ten modifiable risk factors. We used the Logistic procedure in SAS, an example of which can be found in Appendix 3. These models included the predictor variables age (continuous), gender (male, female), and education (did not complete high school, completed high school), and smoking status (current, former, never). Language (English, French) was considered for inclusion[**] but because there were only 4 French speakers, inclusion of this variable yielded incomplete data and so the language variable was excluded.

Our modeling efforts using the dataset under consideration suffered from the small number of observations and the fact that relatively few patients did not have knowledge of most individual risk factors. Table 5 summarizes deficiencies in the data relative to modeling each of the risk factors, as well as the resulting restrictions to separate models used for each risk factor.

[**] The UOHI serves patients in both official languages and so there was interest in whether there were different needs in these two groups.

First, because of the few non-anglophone patients (4 out of 71) we were unable to include the language variable in the models. In most cases, all patients without knowledge of a given risk factor were anglophone and so we are unable to estimate any effect of language on knowledge.

Second, smoking status was an issue in many models. This variable has three levels (never smoked, former smoker, current smoker) and is a typical and clinically meaningful parameterization of this variable. In this study, there were relatively few patients who had never smoked (16 out of 71) and we often had risk factors for which there were no "never smokers" among those unaware of the risk factor. This meant that smoking status could not be included in the model unless we were to collapse the variable into a binary one. While the most straightforward way might seem to create a single level of never smokers and former smokers, we know that some of the former smokers had quit very shortly before admission to hospital oras a result of their diagnosis, making the combination of these levels meaningless. Therefore, we elected to keep the variable parameterization as initially defined and limit the possible models that could be estimated from the data.

Table 5 - Incomplete data issues
For all variables, "1"=aware of RF and "0"=not aware of RF

Independent Variable	*Issue*	*Resolution*
High Cholesterol	All "0"s correspond to English speakers	This is an issue with most of the risk factors. We elect to drop Language from all analysis
	No "0"s are current smokers	Exclude smoking status from model
Hypertension	All "0"s are former smokers	Exclude smoking status from model
Smoking	All "0"s are male, former smokers, HS+ education	Exclude gender, smoking status, and education from model
High Blood Sugar	No "0"s have never smoked	Exclude smoking status from model.
Overweight	No "0"s have never smoked	Exclude smoking status from model
	All "0"s had less than high school	Exclude education from model.
	-	-
	No "0"s have never smoked	Exclude smoking status from model
Inactivity	All "0"s are male	Exclude gender from model.
High Salt Diet	-	-
High Fat Diet	No "0"s have never smoked.	Exclude smoking status from model.

Results for single variable models are presented in Table 6. All variables consume one degree of freedom with the exception of smoking status which consumes two (as it has three levels).

Table 6 - Single Variable Model Results

Risk Factor	Observed Awareness	Model	-2lnL	G
High Cholesterol	68/71	Intercept	24.9	
		Int + Age	24.8	0.1
		Int + Gender	24.8	0.0
		Int + Education	24.1	0.8
Hypertension	69/71	Intercept	18.2	
		Int + Age	17.8	0.4
		Int + Gender	18.1	0.1
		Int + Education	18.2	0.0
Smoking	68/71	Intercept	24.9	
		Int + Age	24.2	0.7
High Blood Sugar	59/71	Intercept	64.5	
		Int + Age	64.2	0.3
		Int + Gender	61.4	3.1
		Int + Education	64.2	0.3
Overweight	68/71	Intercept	24.9	
		Int + Age	24.8	0.1
		Int + Gender	24.8	0.1
Inactivity	61/71	Intercept	57.7	
		Int + Age	55.4	2.3
		Int + Gender	56.0	1.7
		Int + Education	54.1	3.6
		Int+ Smoking Status	52.5	5.2
High Salt Diet	61/71	Intercept	57.7	
		Int + Age	57.1	0.6
		Int + Education	54.1	3.6

Risk Factor	Observed Awareness	Model	-2lnL	G
Low Fibre Diet	35/71	Intercept	98.4	
		Int + Age	97.0	1.4
		Int + Gender	97.7	0.7
		Int + Education	90.5	7.9**
		Int+ Smoking Status	96.2	2.2
High Fat Diet	66/71	Intercept	30.7	
		Int + Age	29.6	1.1
		Int + Gender	30.3	0.4
		Int + Education	30.5	0.2
Stress	65/71	Intercept	41.1	
		Int + Age	41.1	0.1
		Int + Gender	39.7	1.4
		Int + Education	39.5	1.6
		Int+ Smoking Status	40.1	1.0

** Significant at $p<0.005$

This exercise pointed out the relative sparseness of the dataset as there were many problems with incomplete data. This is mainly due to the high knowledge level exhibited for most of the risk factors, yielding only a few patients (often 2 or 3 out of 71) who were unaware of the risk factor. All but 1 of the 10 risk factors were best modeled with an intercept term only, yielding no more information than what was gleaned from the descriptive analysis.

Full results are presented for two of the risk factors. Knowledge of high cholesterol was one of the risk factors with high overall awareness, and the modeling results demonstrate the issues associated with modeling this high probability event (Table 7). Because only 3 of the 71 observations corresponded to the negative response (no knowledge of RF), we were not able to include smoking status in the models because none of these responses included current smokers. Similarly, no interaction terms involving two classification variables could beincluded in the models for analogous reasons.

Table 7: Modeling results for knowledge of high cholesterol as a risk factor for heart disease.

Model	-2 log L	G	Δdf	*p*
Intercept	24.86			
Int + Edu	24.10	0.76	1	*0.38*
Int + Age	24.76	0.09	1	*0.77*
Int + Gender	24.83	0.03	1	*0.86*
Int + SS	*Quasi-complete separation of data points. No current smokers in response=0 group.*			
Int + Edu + Age	24.01	0.85	2	*0.65*
Int + Edu + Age + Edu*Age	21.22	3.63	3	*0.30*
Int + Edu + Gen	24.09	0.77	2	*0.68*
Int + Edu + Gen + Edu*Gen	*Quasi-complete separation of data points. No female with high school or greater education in response=0 group.*			
Int + Age+ Gen	24.75	0.11	2	*0.95*
Int + Age + Gen + Age*Gen	*Quasi-complete separation of data points*			
Int + Edu + SS	*Quasi-complete separation of data points*			
Int + Edu + SS + Edu*SS	*Quasi-complete separation of data points.*			
Int + Edu + Age + Gender	24.00	0.85	3	*0.84*
Int + Edu + Age + Gender + SS	*Quasi-complete separation of data points.*			

We also provide modeling results for the risk factor "low fibre diet", for which there was much lower awareness (35 out of 71 patients). Table 8 provides results for these models. We note first that education is the only individual variable to provide significant improvements in fit to the data. Adding age, gender, or smoking status does not significantly improve the model fit, nor does including an interaction term.

Table 8 - Modeling results for knowledge of low fibre diet as a risk factor for heart disease

Model	-2 log L	G	Δdf	p	Comments
Intercept	98.41				
Int + Edu	90.50	7.91	1	0.005	Designate as (*)
Int + Age	97.03	1.38	1	0.24	
Int + Gender	97.73	0.68	1	0.41	
Int + SS	96.23	2.18	2	0.34	
Int + Edu + Age	88.87	9.55	2	0.009	Addition of age to (*) not significant improvement in -2logL (Δ=1.63, 1 df)
Int + Edu + Age + Edu*Age	88.86	9.55	3	0.02	Add'n of interaction not a significant improvement in -2logL (Δ=0.01, 1 df))
Int + Edu + Gen	90.10	8.31	2	0.02	Addition of gender to (*) not significant improvement in -2logL (Δ=0.40, 1 df)
Int + Edu + Gen + Edu*Gen	87.56	10.85	3	0.01	Add'n of interaction not a significant improvement in -2logL
Int + Age+ Gen	96.52	1.90	2	0.39	
Int + Age + Gen + Age*Gen	95.85	2.57	3	0.46	
Int + Edu + SS	88.90	9.51	3	0.02	Add'n of smoking status to (*) not significant improvement in -2logL (Δ=0.460, 2 df)
Int + Edu + SS + Edu*SS	87.92	10.50	5	0.06	
Int + Gen + SS	95.54	2.87	3	0.41	
Int + Gen + SS + Gen*SS	95.03	3.38	5	0.64	
Int + Age + SS	93.91	4.50	3	0.21	
Int + Age + SS + Age*SS	92.64	5.78	5	0.33	
Int + Edu + Age + Gender	88.61	9.80	3	0.02	

We thus conclude that the best model for the low fibre diet risk factor is an intercept term and the variable education. This model takes the form

$$\log\left[\frac{p(awareness)}{1-p(awareness)}\right] = 0.55 - 1.39(edu2)$$

where edu2 takes the value 1 if the patient does not have a high school diploma and 0 for high school or greater. The odds ratio was 0.25 with a 95% confidence interval of (0.09, 0.68). While the overall knowledge probability for this risk factor was 0.49, 63% of patients with at least a high school education were aware of this RF vs. 30% of those with less than high school.

Table 9 summarizes the best models.

Table 9 – Best Models

Risk Factor	Observed awareness	Fitted model
Hypertension	0.97	Intercept
High Cholesterol	0.96	Intercept
Smoking	0.96	Intercept
Overweight	0.96	Intercept
High Fat Diet	0.94	Intercept
Stress	0.92	Intercept
Inactivity	0.86	Intercept
High Salt Diet	0.86	Intercept
High Blood Sugar	0.83	Intercept
Low Fibre Diet	0.49	Int + Education

Given the results of the descriptive analysis in the previous section, where the middle age group (50s and 60s) had the highest overall knowledge score, we swapped the continuous age variable in the above model with the three-category variable used previously. The continuous variable would only permit a monotonic relationship to be modeled and we thought that perhaps a U-shaped curve would improve the fit, at the expense of one degree of freedom. None of the models were found to improve with this modification.

Further exploration of these models suggested that one case was exerting a high degree of influence. We determined this by examination of the impact of the removal of individual observations on the parameter estimates, on the deviance and chi-squared statistics, as well as on the Pearson and deviance residuals. One case yielded high residuals, had a large impact on the deviance and chi-square statistics, as well as high impact on some of the parameter estimates. This observation corresponded to an 81-year-old male, former smoker, with a high school education, who answered most of the risk factor knowledge questions with "Don't Know". We removed this observation from the dataset and refit those models for which the probability of a correct response for the risk factor was equal to or less than 0.90 (this avoided modeling data with only one or two data points for which the response variable was "'0'). Based on this dataset of 70 patients, new models were chosen for Inactivity and High Salt Diet.

The model fitting process for Inactivity is summarized in Table 10. Based on these models, we selected education and smoking status as the variables to include in this model.

Table 10 - Model for Inactivity with influential data point Removed (n=70)

Model	**-2lnL**	**G**
Intercept	53.7	
Int+age	52.4	1.3
Int+Gender	52.5	1.2
Int+Education(Edu)	48.5	5.2*
Int+Smoking Status (SS)	46.8	6.9*
Int+SS+Age	43.0	3.8
Int+Edu	48.5	5.2*
Int+Edu+SS	42.3	6.2*
Int+Edu+SS+Age	38.8	3.5
Int+Edu+SS+Gender	41.7	0.6

A similar exercise was conducted for High Salt Diet. Because of the problems with incomplete data, it was only possible to fit models including age and education. The model including education resulted in a change in deviance of 5.235 ($p<0.025$) but no further improvement was obtained by including age.

Models for High Blood Sugar and Stress were not improved with removal of the influential point. Table 11 summarizes the changes to the models using the reduced dataset.

Table 11 - Comparison of selected models before and after removal of one influential data point.

Risk Factor	Observed awareness	Model using all data	Model excluding influential data point
Hypertension	0.97	Intercept	Intercept
High Cholesterol	0.96	Intercept	Intercept
Smoking	0.96	Intercept	Intercept
Overweight	0.96	Intercept	Intercept
High Fat Diet	0.94	Intercept	Intercept
Stress	0.92	Intercept	Intercept
Inactivity	0.86	Intercept	Int + Education + Smoking Status
High Salt Diet	0.86	Intercept	Int + Education
High Blood Sugar	0.83	Intercept	Intercept
Low Fibre Diet	0.49	Int + Education	Int + Education

5.2.4. Discussion

Of the 10 modifiable risk factors, we were able to find significant relationships to one or more of the explanatory variables in three cases: low fibre diet, high salt diet, and inactivity. Education was found to significantly contribute in each of these models and was related to the risk factor as expected: those with high school or greater education were more likely to be aware of the risk factor. For inactivity, smoking status was also significantly related to knowledge. Relative to those who had never smoked, current smokers were less likely to be aware of this risk factor and

former smokers were more likely to be aware. The predicted probabilities are displayed inTable 12.

Table 12 - Predicted probabilities for knowledge of inactivity as a risk factor by education and smoking status

Smoking Status	No High School	HS Grad
Current	0.53	0.86
Former	0.91	0.98
Never	0.64	0.91

Awareness is high (over 90%) for former smokers regardless of education, but for current and never smokers, education makes a considerable difference in their knowledge levels. These probabilities suggest a hypothesis that former smokers have been the recipient of educationabout risk factors as evidenced by having stopped smoking, and hence are more likely to be aware of other risk factors (such as inactivity).

The logistic regression modeling was hampered by a couple of key issues. The first is the relatively few patients who were unaware of certain of the risk factors. While a positive finding from a public health point of view, attempting to understand what kind of people are unaware of a risk factor when only 3 patients out of 71 are in that category is close to futile.

The second issue is that of incomplete data. Seven of the 10 risk factors had such problems, excluding one or more variables from consideration in the models. Smoking status had to be excluded from seven of the ten models mainly due to the lack of individuals who had never smoked and who were unaware of the risk factor. Consideration was given to collapsing the

smoking status variable, but there is no a priori best way to do this. The measurement of smoking status in the three categories used here (current, former, and never smoked) is quite standard in the literature and represent well understood subgroups of the population.

Both of these issues (high observed probabilities and incomplete data) could be addressed by larger sample sizes and a different sampling strategy. This will be addressed in Section 5.5.

Finally, we must note that we have not corrected the Type II error to account for multiple comparisons.

5.3. Correspondence Analysis (CA)

5.3.1. Introduction to CA

CA is a data visualization technique that reduces a multidimensional dataset to a smaller number of dimensions so that relationships can be more easily visualized. It is a descriptive method similar in character to construction of a histogram to view the key features of a dataset. In particular, CA examines the relationship between the rows and columns of a matrix of cross-classified data and describes how they deviate from independence.

For example, from the dataset at hand, we can examine the data on smokingstatus and education depicted in Table 13.

Table 13 - UOHI data on smoking status and education

	Less than High School	*High School Grad*	*College Grad*	*University Grad*	*Total*
Current Smoker	7	1	4	5	12
Former Smoker	17	12	7	7	43
Never Smoked	6	4	2	4	16
Total	30	17	13	11	71

Under independence, we would expect the following distribution of observations (Table 14).

Table 14 - Expected values of data in previous table under independence of rows and columns

	Less than High School	*High School Grad*	*College Grad*	*University Grad*	*Total*
Current Smoker	5.1	2.9	2.2	1.9	12
Former Smoker	18.2	10.3	7.9	6.6	43
Never Smoked	6.7	3.8	2.9	2.5	16
Total	30	17	13	11	71

CA explores how the observed data deviate from that expected under independence, using the row and column categories that account for the greatest proportion of the deviation.

The data decomposition process and associated notation are outlined in Appendix 5, using the example above. The remainder of this section describes the CA visualization process and results for the dataset at hand. In brief, by using eigenvectors, every row and column profile can be thought of as a point in space. Points that are relatively close together can be thought of as similar, and points that are far apart as not similar. Viewing these points in 2- or 3- dimensional space can often permit us to see these distances. Visualizing the separation of points is enhanced by choosing the axes in the reduced space that lie closest to all the points. The point where these axes cross is called the centroid and is the weighted average of the row profiles. The variance is a measure of the spread of these points around the centroid, the weighted chi-square distance where the weights are the row masses.

The CA literature uses the concept of "inertia" (Λ^2) which is related to the chi-square statistic (X^2) as

$$X^2=\Lambda^2 N$$

where N is the number of observations in the table. It is equivalent to Pearson's contingency coefficient (ϑ^2)[19], the deviation from independence of a contingency table.

The inertia can be decomposed into eigenvalues. Eigenvalues (or the squares of the singular values) correspond to the amount of inertia explained by each of the dimensions. We can express the amount of inertia explained by each dimension as a percentof total inertia. In viewing a decomposition in two dimensions, we are looking at the inertia explained by the dimensions that explain the largest two values of inertia. By summing these together, we can see how much of the variability in the problem we are able to see in a two-dimensional graphic.

In our example (Table 13), we obtain the following decomposition:

Principal Inertia	Percent	Cumulative Percent
0.6535	26.14	26.14
0.5372	21.49	47.63
0.5000	20.00	67.73
0.4628	18.51	86.14
0.3465	13.86	100.00

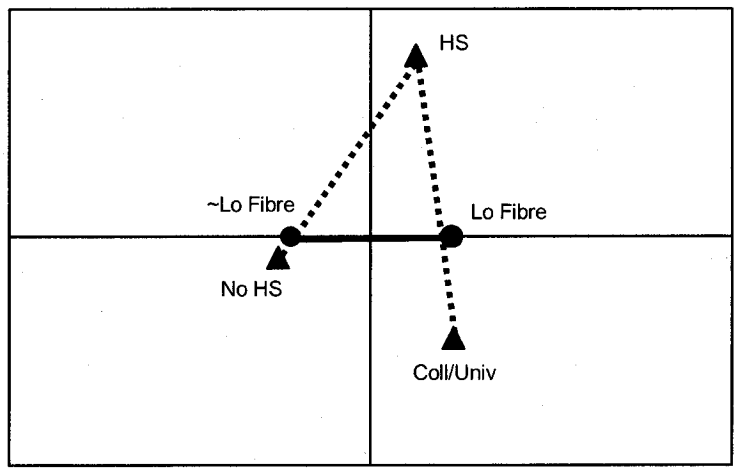
In this example, there is not very much difference in the percent of inertia contributed by each dimension, making the 2-dimensional reduction for visualization less useful than it might be. The largest two dimensions only account for under 48% of the total inertia.

5.3.2. Results from Simple Correspondence Analysis

We elected to start with simple analyses looking at some of the individual risk factors and the education variable that appeared in some of the logistic regression models. In particular, we completed the analysis for low fibre diet, high salt diet, and inactivityfor which the logistic regression models included education as a significant predictor variable. We also looked at high cholesterol diet, for which modeling yielded no predictor variables of significance, and which had high patient knowledge.

The following figures depict results for each of the first three risk factors separately: low fibre diet (Figure 8), high salt diet (Figure 9), and inactivity (Figure 10).

Figure 8 - CA Plot: Low Fibre Diet and Education
Inertia=86.3%



The labels NoHS, HS, and Coll/Univ refer to the three levels of education. Risk factor knowledge is labeled with the risk factor name, preceded by a tilde (~) for lack of knowledge (e.g., Lo Fibre vs. ~Lo Fibre)

The point representing lack of knowledge of low fibre diet lies close to the point representing the attribute of no high school education, as would be expected given thelogistic modeling results. Similarly, knowledge of high salt diet (Figure 9) and of inactivity (Figure 10) as risk factors are also close to no high school education, although not as markedly so as in the case of low fibre diet. (Percents of inertia represented in the two dimensional representation were 86.3%, 74.3%, and 74.7% respectively.)

Figure 9 - CA Plot for High Salt Diet and Education
Inertia=74.3%

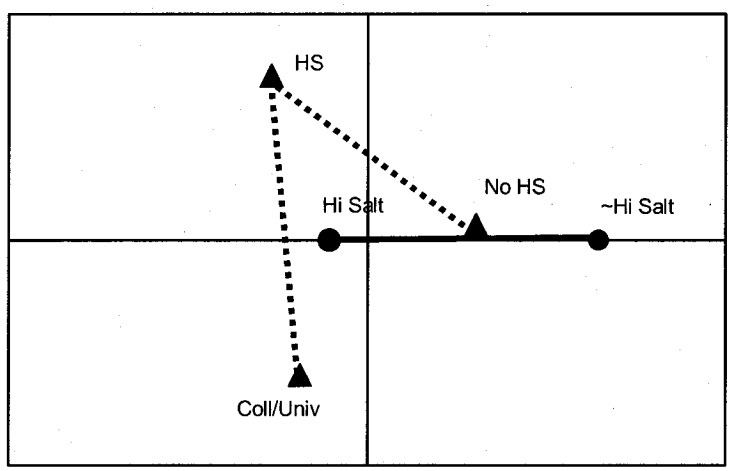
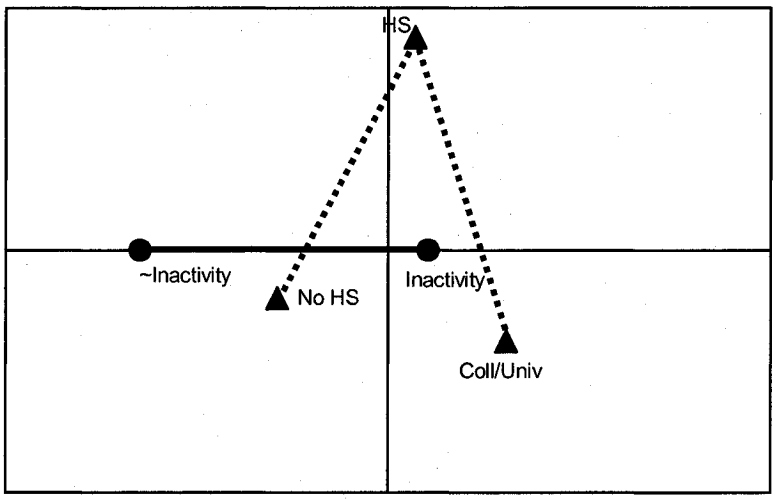
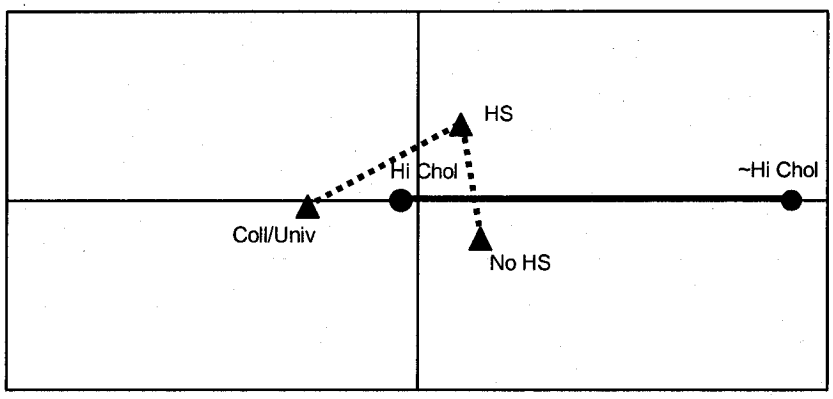


Figure 10 – CA Plot: Inactivity and Education
Inertia=74.7%



The CA plot for high cholesterol (Figure 11) is of note as the point representing lack of knowledge is at a great distance from any of the other points. While itis the furthest from College/University education, we note that those who are aware of high cholesterol as a risk factor are equally close to all three educational levels. This would suggest that education has a less important role in knowledge of this risk factor. (76.3% of the inertia is captured in the two-dimensional representation for this risk factor.)

Figure 11 - CA Plot: High Cholesterol and Education
Inertia=76.3%



The next step was to consider more than one risk factor in the CA. We began by examining low fibre diet and high salt diet with education, for which 63% of the inertia was captured in two dimensions.

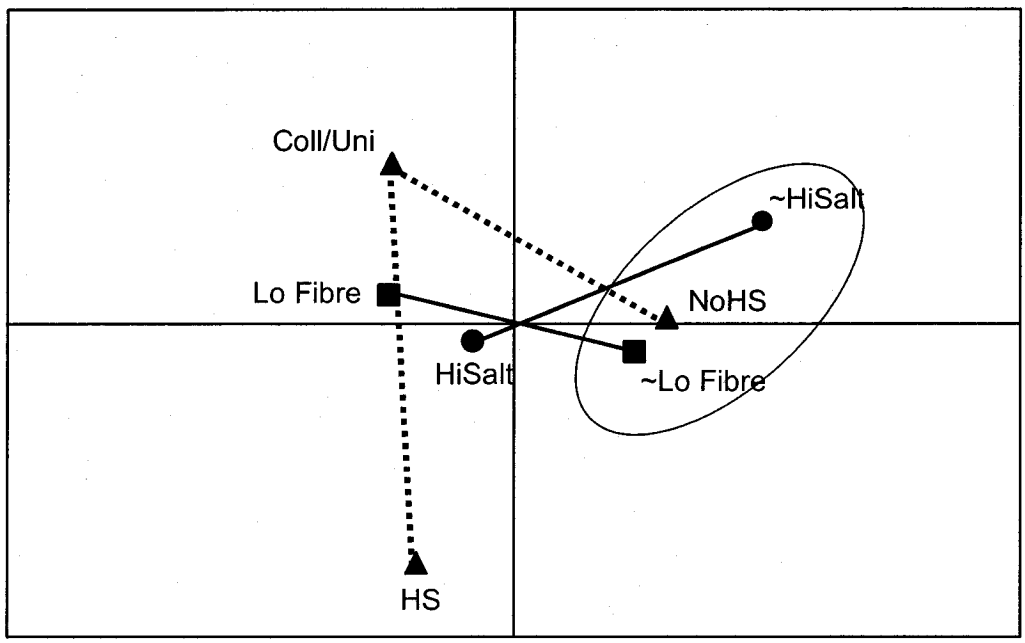
We used multiple correspondence analysis (MCA) to examine the bivariate relationships between pairs of a set of categorical variables. In this case, data are entered into the analysis using a Burt matrix, a symmetric partitioned matrix of blocks that contain the two-way frequency tables for each pair of variables. For the example under consideration, the data matrixis displayed in Table 15. MCA is a singular value decomposition of this matrix.

Table 15 - Example Burt Matrix for 3 Categorical Variables

	NoHS	HS	Uni	LoFi	~LoFi	HiSalt	~HiSalt
NoHS	30	0	0	9	21	23	7
HS	0	17	0	10	7	16	1
Uni	0	0	24	16	8	22	2
LoFi	9	10	16	35	0	32	3
~LoFi	21	7	8	0	36	29	7
HiSalt	23	16	22	32	29	61	0
~HiSalt	7	1	2	3	7	0	10

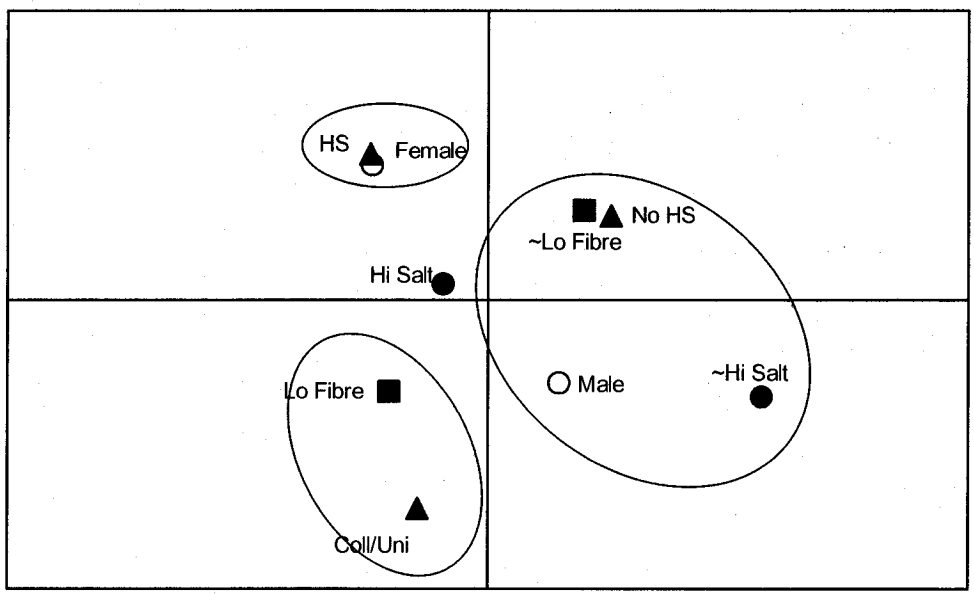
Again, we see that lack of knowledge of these two risk factors is associated with not completing high school (Figure 12).

Figure 12 -CA Plot: Low Fibre Diet, High Salt Diet and Education
Inertia=62.6%



We now elected to add in another possible "explanatory" variable: gender. This increases the dimensionality of the problem and therefore the percent inertia capturedin two dimensions drops to 55.7%. The results are displayed in Figure 13. We note here that lack of knowledge of the risk factors "high salt diet" and "low fibre diet" cluster together with low education (not a high school graduate).

Figure 13 - CA Plot: Low Fibre, High Salt, Gender and Education
Inertia=55.7%

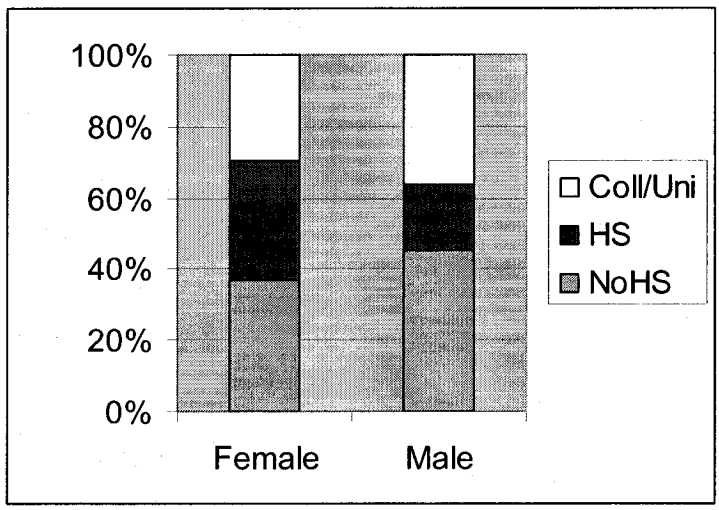


The attributes female and high school education are almost coincident, while the attribute male is almost equidistant from no high school and college/university grad. Knowledge of a low fibre diet and gender are orthogonal, underscoring the lack of relationship that was suggested in the logistic modeling exercise.

The coincidence of points for high school education and female may be due to the relative distribution of education by gender. Figure 14 demonstrates that the distribution of education

level categories for women is fairly even, whereas men were more likely to have either college/university level education or to not have finished high school. This is suggestive only; analysis of the two-way contingency table of education by gender does not reject the hypothesis of independence of these variables. However, the figure suggests that females are more likely to have a high school education than males (33% of women vs. 18% of men).

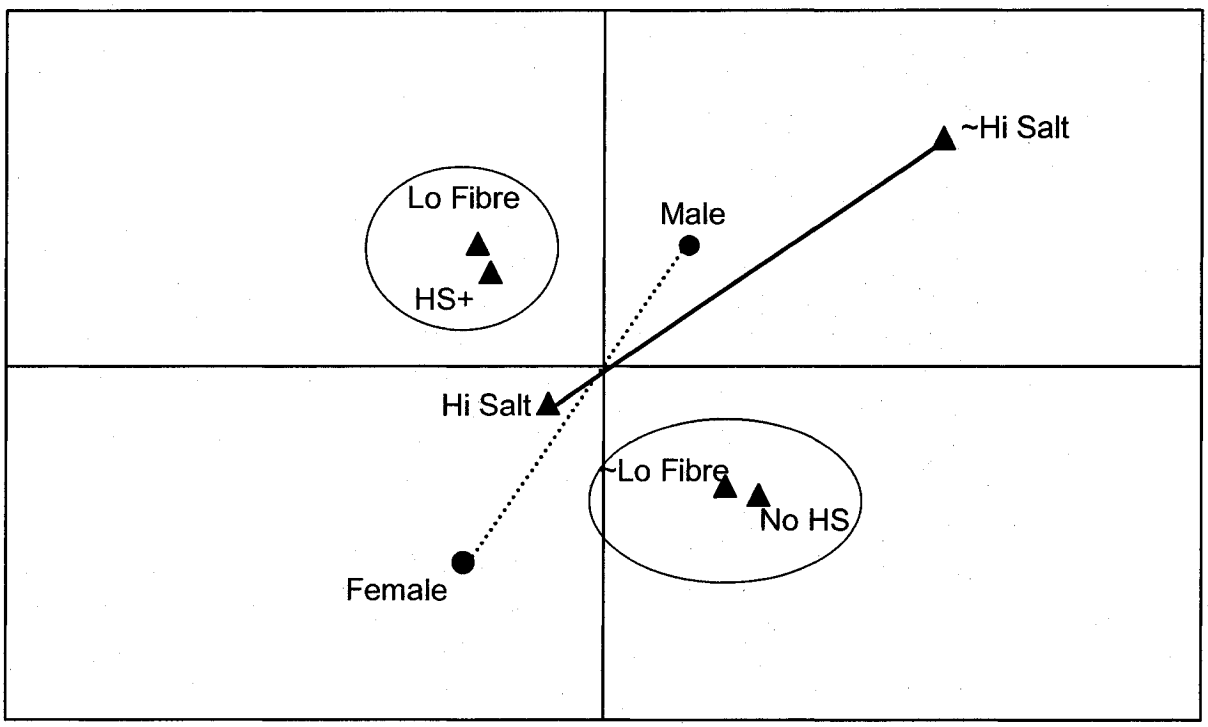
Figure 14 - Distribution of education by gender



Previously (section 4.2) we noted that risk factor knowledge scores for high school grads and college/university grads did not differ significantly. To simplify the plots and reduce the dimensionality of the problem, we elected to look at a binary variable for education (No HS, HS+). We repeated the previous analysis including high salt diet, low fiber diet, gender and this new binary education variable (Figure 15).

We can see that education and knowledge of a low fibre det are very related, whereas gender and knowledge of high salt diet are to a great extent orthogonal to them.

Figure 15 - CA Plot: Low Fibre Diet, High Salt Diet, Gender and Education(2)



Next we add in knowledge of stress as a risk factor. We obtain very similar results to those above, with the stress axis lying very close to the high salt axis (Figure 16). We note four clusters, one in each of the quadrants, as follows:

(1) ~High Salt, ~Stress

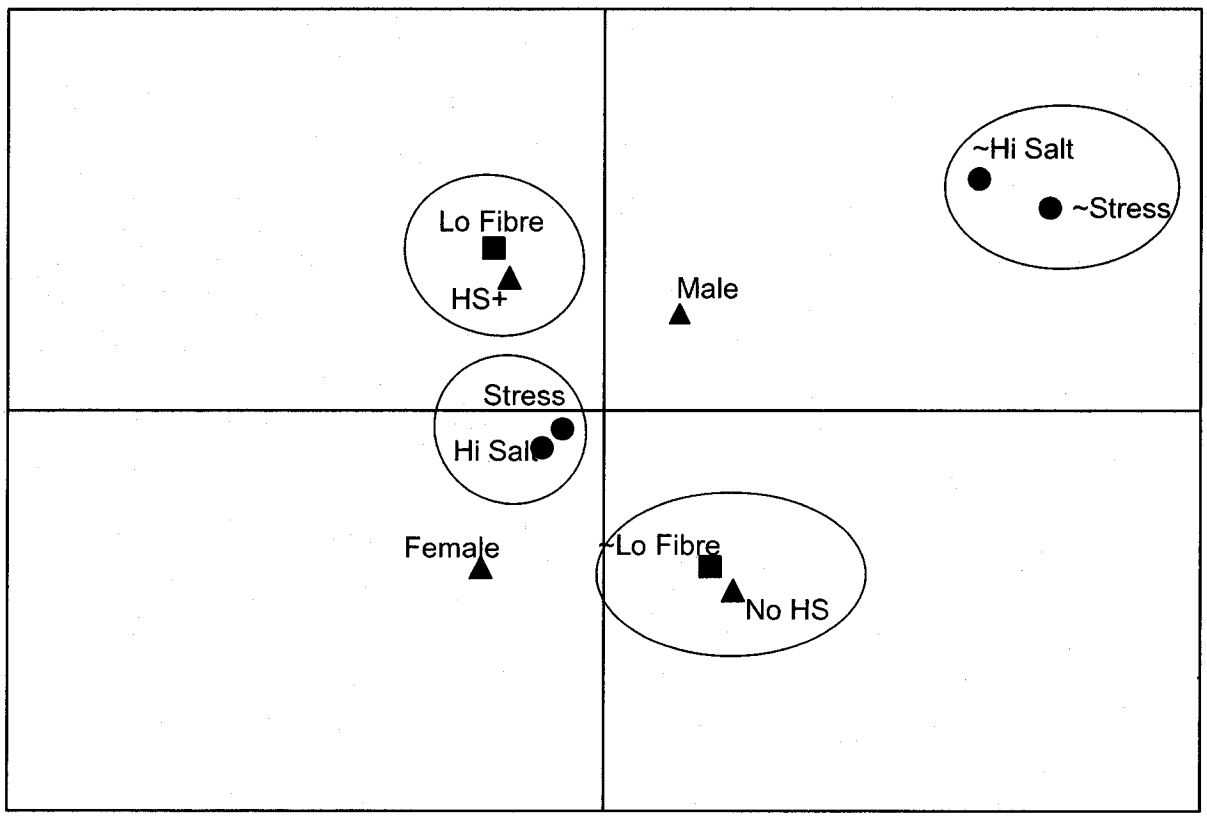
(2) No HS, ~Low Fibre Diet

(3) Stress, High Salt Diet

(4) High School or greater, Lo Fibre Diet

As in the previous analysis using the three-level education variable, knowledge of a low fibre diet tends to vary by education. We note that being male lies midway between knowledge and lack of knowledge of all three risk factors, suggesting that there is not a strong gender component here.

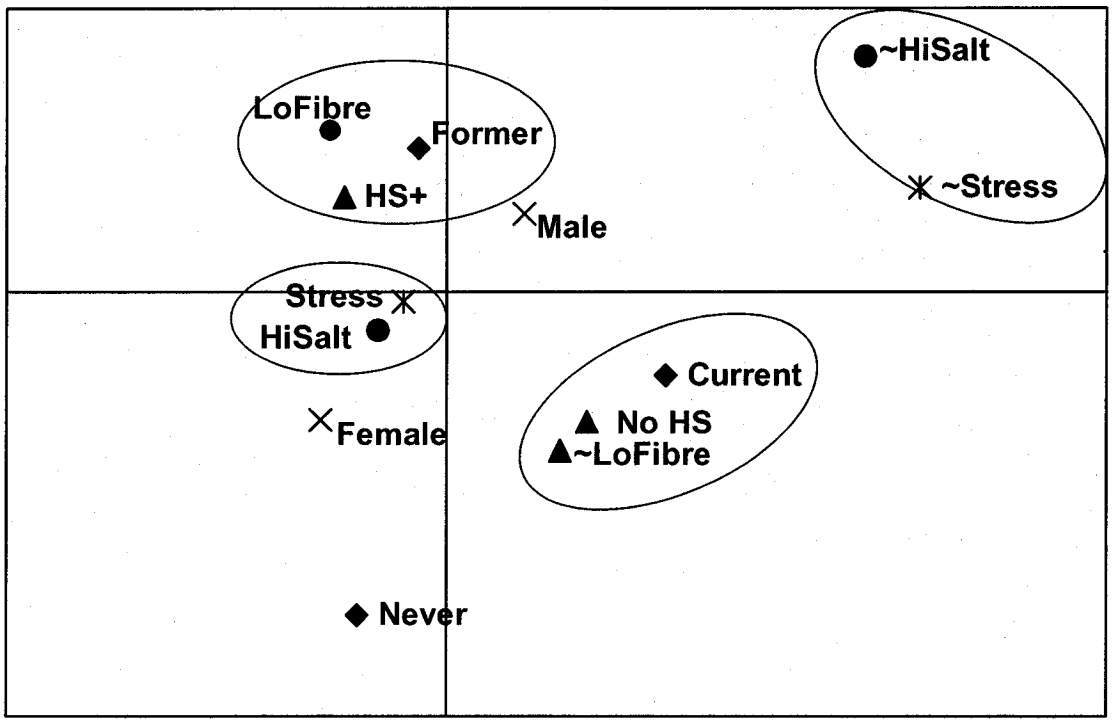
Figure 16 - CA Plot: High Salt Diet, Low Fibre Diet, Stress, Gender, Education(2)



Smoking status (current smoker, former smoker, never smoked) is another variable of interest as it played a role in predicting knowledge of inactivity as a risk factor. Figure 17 presents the analysis as above but including smoking status.

Figure 17 - CA Plot: High Salt Diet, Low Fibre Diet,

Stress, Gender Education(2) and Smoking Status

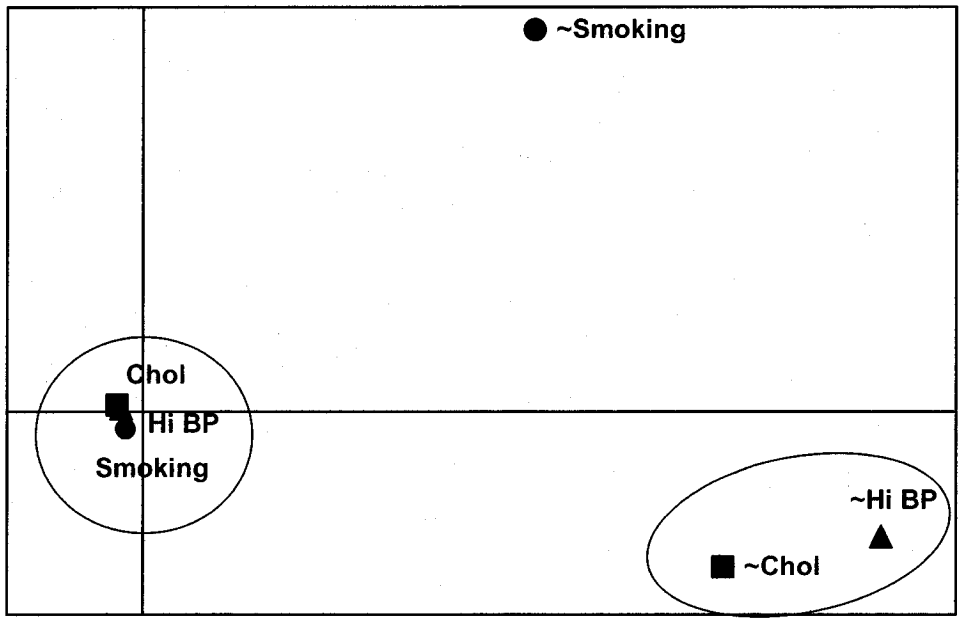


We see similar clusters to those in Figure 16 with some shifting. The attribute of current smoking joins the ~ low fibre diet and no high school cluster. Being a former smoker is associated with knowledge of low fibre diet and high school or greater education cluster. Never smokers lie outside of the clusters as previously defined.

Also of interest is knowledge of the top three risk factors: high cholesterol, smoking, and hypertension. We will use CA to understand how knowledge of these risk factors relates to patient characteristics.

We start with the three risk factors alone. Figure 18 displays the CA plot for these three factors. We note that points corresponding to lack of knowledge lie well away from the centroid, close to which lie the points representing knowledge. Points representing lack of awareness of cholesterol and hypertension lie close together, while lack of awareness of smoking is a lone point in its quadrant. This plot represents 94% of the inertia, for two of the three dimensions associated with the three risk factor representation.

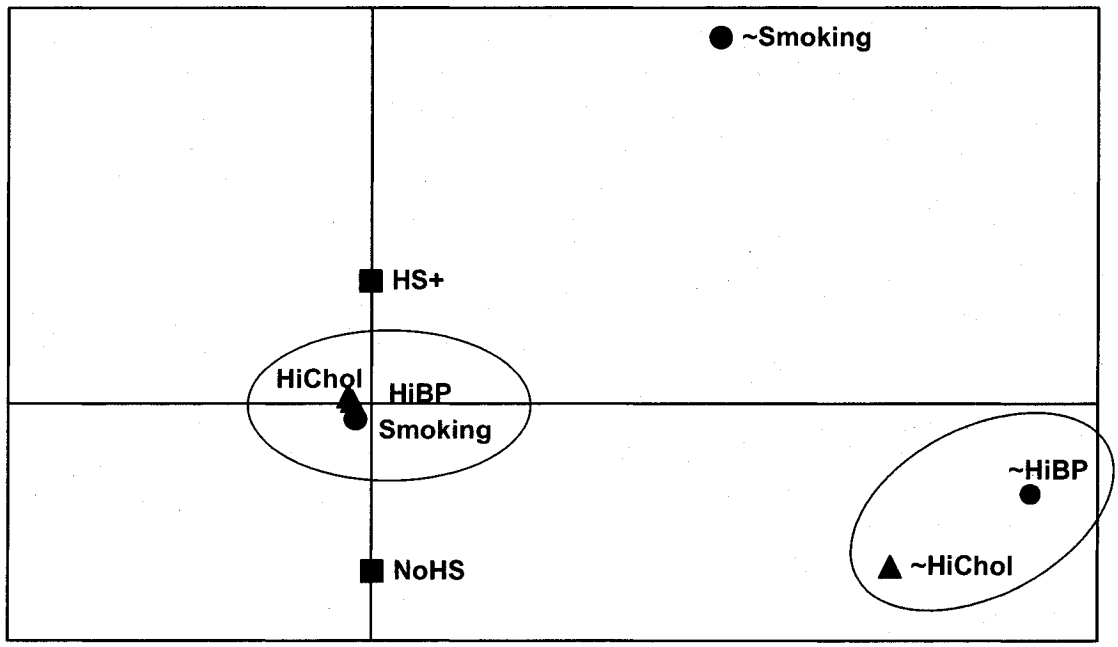
Figure 18 – CA Plot: Knowledge of High Blood Pressure, High Cholesterol, and Smoking
Inertia=93.9%



Now we add education to this analysis, which yields Figure 19. We note that knowledge of the top three risk factors lies midway between the two education attributes, suggesting that education does not play a significant role for those who are aware of these factors. Lack of awareness of smoking as a risk factor lies closer to the attribute of higher education, while lack of awareness of

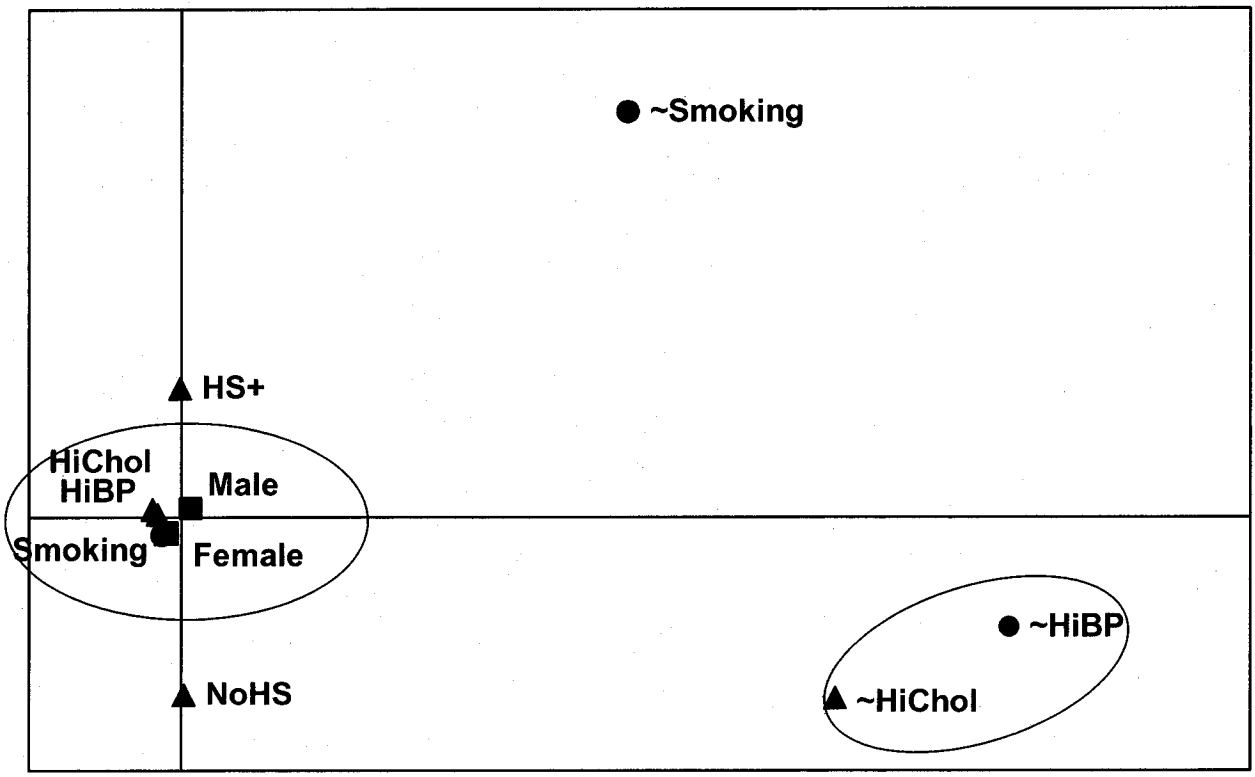
high blood pressure and high cholesterol lie closer to the attribute of no high school diploma. However, these are weak associations at best.

Figure 19 - CA Plot: Top 3 Risk Factors and Education(2)
Inertia=79.1%



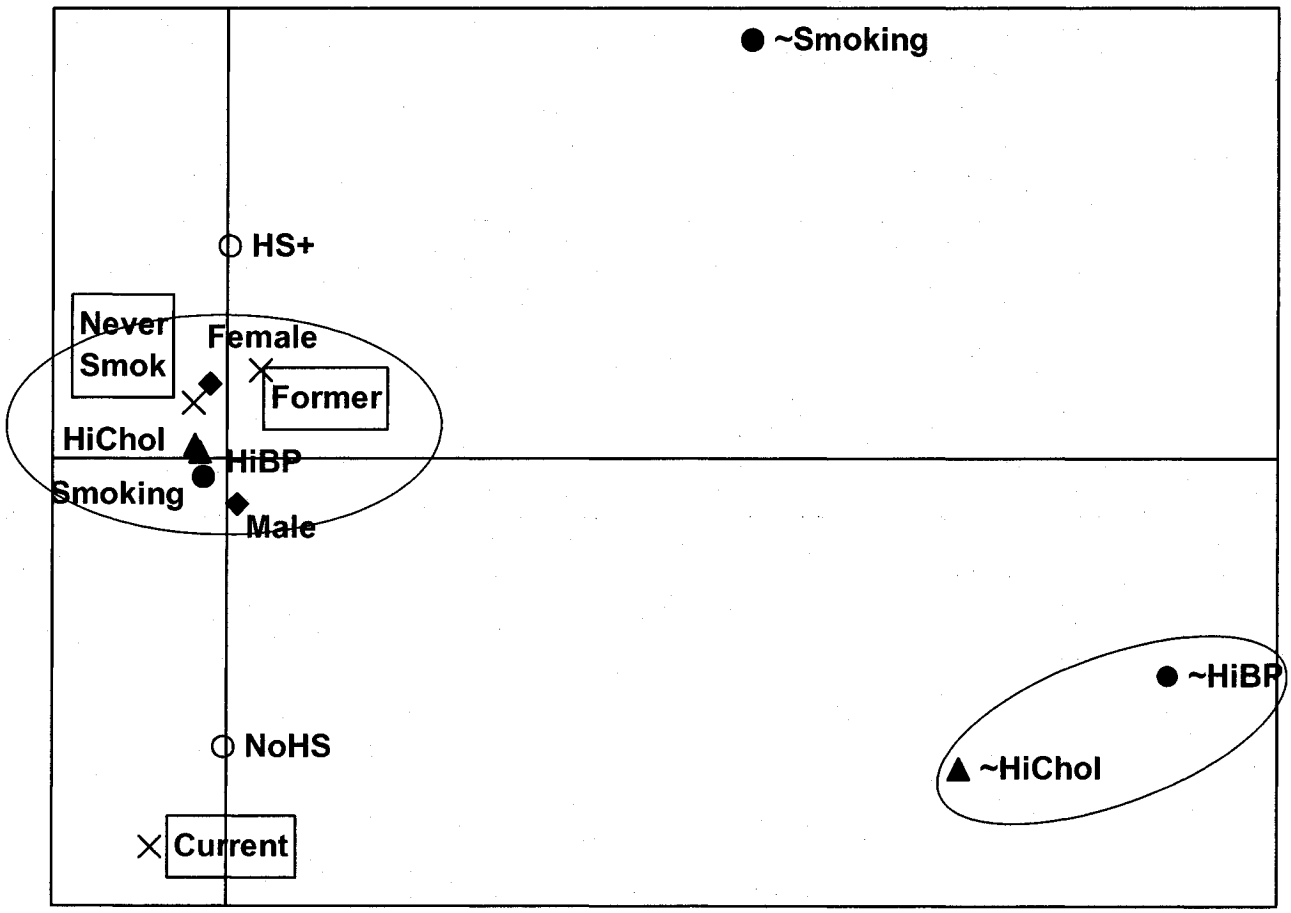
Next, we include gender in the analysis (Figure 20). We note that both gender attributes are clustered closely at the centroid, suggesting that gender is not a strong predictor of knowledge of these top three factors.

Figure 20 - CA Plot: Top three risk factors with education(2) and gender
Inertia=64.4%



Finally, we add smoking status to the analysis. We see in Figure 21 that while former smokers and never smokers are clustered around the centroid, current smokers are in thelower left quadrant, close to the attribute of no high school diploma. They are also closer to the lack of knowledge of high cholesterol and high blood pressure. Current smokers are located further from lack of knowledge of smoking as a risk factor than from the point representing knowledge of smoking suggesting that they are very aware of the risks.

Figure 21 - CA Plot: Top three risk factors with
gender, education, and smoking status.
Inertia=47.1%



5.3.3. Discussion

We have looked at a subset of the risk factors, some with high awareness levels and some with lower levels. This section is designed to demonstrate the possibilities of correspondence analysis and not an exhaustive study of all ten risk factors and the possible explanatory variables. The power of this technique to provide at least directional findings is clear and we put forward the following conclusions:

1. Knowledge of the top three risk factors (smoking, high cholesterol, hypertension) is not obviously related to gender or education.

2. Those who are unaware of smoking as a risk factor are not current smokers.

3. Knowledge of low fibre diet is related to education, but not to gender.

4. Females are more aware of high salt diet and stress as cardiac risk factors, and are more likely to have never smoked.

5. Smokers tend to have lower education and be unaware of the risk of a low fibre diet.

These types of statements are suggested by the analysis of this section, but the nature of CA does not permit quantification of the relationships.

Overall, the high awareness risk factors (for example, smoking, high blood pressure, high cholesterol) were not able to be related to specific explanatory variables, whereas certain risk factors with lower awareness (for example., low fibre diet, low salt diet, stress) were more likely to cluster with certain demographics.

5.4. Comparison of the two approaches

5.4.1. General

The goal for this thesis was to consider two different approaches to the analysis of multidimensional data. Logistic regression was hampered by sample size and high levels of awareness, resulting in quantifiable parameter estimates for only three of the ten risk factors, two of which were achieved after removal of an influential data point. Table 16 summarizes the findings of the logistic regression approach.

Table 16 - Logistic Regression Results - Summary

Risk Factor	Explanatory variables	Relationship
Low Fibre Diet	Education	No HS: 30% awareness HS+: 63% awareness
Inactivity	Education Smoking Status	See Table 12
High Salt Diet	Education	No HS: 77% awareness HS+: 95% awareness

There is a well-established literature in logistic regression regarding assessment of influential data points and this was helpful for this exercise.

Correspondence analysis has fewer constraints regarding sparse data, but does not yield quantifiable relationships. For the risk factors we examined in this paper, we were able to see some relationships with gender, smoking status, and education that were not visible using logistic regression. These are summarized in section 5.3.3. The technique one uses in a research study of this nature will depend on the intended use of the results.

5.4.2. Combining these approaches

In observing the clustering of certain risk factors in the CA, we decided to combine the two techniques as follows. In Figures 17 & 18, we see that knowledge of the risks of stress and a high salt diet tend to cluster together, meaning that they tend to occur in the same patients.

We created a new variable that takes the value "1" if the patient has knowledge of both risk factors and "0" otherwise. The observed frequency of the new variable is 83% (that is, 83% of patients have knowledge of both risk factors) vs. 92% for stress and 86% for high salt diet.

Assessing potential predictors as in Section 5.2 yields a "best model" that includes education and gender. The models yields the following predicted probabilities of knowledge of both risk factors (Table 17).

Table 17 - Predicted probabilities of knowledge of high salt diet and stress, by gender and education

	No HS	HS+
Male	0.59	0.89
Female	0.93	0.99

By using information gleaned from the CA, we are able to improve our ability to predict knowledge using the logistic regression approach.

5.4.3. Usefulness of results

Quantifiable results are important and necessary for certain types of applications. Policy and regulatory decision-making rely on numerical results to compare against industry standards, benchmarks, regulatory requirements, and for forecasting demand. For example, in pharmaceutical studies, numerical results are critical for comparing the efficacy of a new drug to existing drugs or the current standard of care. Regulatory agencies also require precise information about adverse events. Marketing groups want to be able to predict demand for a new device or medication based on measurements of patient and/or clinician preferences. Results of such studies are often directly used in forecasting models to estimate demand at, say, different price points (relative to existing products).

On the other hand, qualitative results are often more useful in exploratory studies, early on in the research process. Prior to heavy investment in a quantitative study, researchers may wish to explore possible hypotheses, relationships, even study designs, and are willing to forgo numerical results for those of a more directional nature. As well, there are arenas in which qualitative results are suitable and sufficient. Early concepts for new products are typically tested using qualitative techniques, as are many advertising campaigns and in the evaluation of product and company names and slogans. In health care, product packaging, information leaflets, office and hospital design, and new medication delivery methods might all be researched using primarily qualitative methods. These types of techniques can also be used in analysis of data that was collected for another purpose and may not have the sample size for more rigorous analysis of relationships of interest. In particular, data from national health and other large surveys and databases can be analyzed post hoc prior to investing in specific studies investigating a particular hypothesis.

In the case at hand, one of the interests of the nursing staff was how to better tailor educational materials to patients in the Heart Institute. The directional results provided by a technique like correspondence analysis can assist them in this task, even in the absence of "hard" numerical data. It enables them to design target audiences for important messages to patients and their families about managing their illness and secondary prevention.

5.4.4. Communication

There are standard ways to present quantitative results that are widely accepted by the scientific community. Tables of numbers, graphs, and model results are examples of methods to communicate numerical results. Logistic regression is a useful and widely used tool to provide these types of results as are descriptive techniques as presented in the journal article proceeding from this research (Appendix 4). Comparison of different studies examining the same or similar

research questions is much easier with numerical results and the translation of research results into information suitable for dissemination to the public and for policy making is much more straightforward. Statements like "Bicycle helmets reduce the chance of death in an accident by x%" or "One in nine women will suffer from breast cancer in their lifetime" result from a series of quantitative studies. In this study, we could say that those with a high school or greater education had more than twice the awareness that a low fibre diet is a risk factor for heart disease, based on the logistic regression results (Table 16).

With correspondence analysis and other techniques with non-numerical output, communication is somewhat more difficult. CA, for example, is not that well known outside of the social science literature. The results need to be presented in a graphical manner, and the interpretation can be very subjective. It is more difficult to communicate results of this type of analysis in words, making publication and dissemination more difficult. Results cannot typically be summarized in a single sentence, or "sound bite". However, as noted above, for many applications, the information does not need to be widely disseminated as it is often for internal use. In this research, it is not the specific results of the study that need to be disseminated. Rather, the results of the study will be used to improve patient-oriented publications.

5.5. Implications for experimental design

Future studies of this type may be undertaken by the UOHI or other institutions. It is therefore of interest how the design of the study might be improved so that, in particular, the descriptive analysis and logistic regression might be more productive.

First, we note that we were working with a small population. We assessed every patient in the UOHI on the data collection day for inclusion in the study, so simply increasing sample size but retaining the rest of the study design would not have been possible. The one day census design was based on a previous study and was helpful for profiling the patient population, one of the objectives of the original study.

For the type of quantitative analysis we undertook in this thesis, a larger sample size would have improved our ability to uncover relationships in the data. There are a couple of ways that this could be accomplished. First, we perform some sample size calculations to understand the magnitude of the issue.

5.5.1. Some sample size calculations

We refer the reader to Chapter 7 of Friedman et al[20] for an excellent review of sample size calculations for different types of data. We consider the situation where we wish to compare the binary response probability in two independent samples. This was of interest in this study, for example, in comparing awareness rates for a risk factor in two sub-populations (say, male and female). We wish to test the hypothesis H_0: $p_1-p_2=0$ where p_1 and p_2 are the response probabilities associated with the two sub-populations. With a large enough sample size, the usual test statistic is

$$Z = (\hat{p}_1 - \hat{p}_2) / \sqrt{\bar{p}(1-\bar{p})\left[\frac{1}{N_1} + \frac{1}{N_2}\right]}$$

where $\hat{p}_1$ and $\hat{p}_2$ are the observed probabilities in the two groups, N_1 and N_2 are the number of observations in each group, and $\bar{p}$ is the pooled observed probability. We reject H_O if Z is greater than the critical value which is determined by the Type I error (α) desired (typically 0.05). For a two-sided test, we reject H_O when $Z>1.96$ for $\alpha=0.05$.

Suppose we wish to be able to observe (statistically) a given difference in observed probability. For example, suppose it is determined that a difference of 20% in awareness of a certain risk factor would be of interest to clinicians and would lead to, say, separate educational materials for the two groups. We would then require sufficient sample size so that a true difference of that magnitude could be observed with a certain power (1-β, the probability of correctly rejecting H_O). Then, the sample size required is:

$$2N = 2\left\{Z_\alpha \sqrt{2\bar{p}(1-\bar{p})} + Z_\beta \sqrt{p_1(1-p_1) + p_2(1-p_2)}\right\}^2 / (p_1 - p_2)^2$$

where there are N subjects in each group for a total of 2N subjects.

We performed a number of calculations for various probabilities and desired observed differences using a power (1-β) of 0.90. Table 18 presents the findings.

Table 18 - Sample Size Calculations
(α=0.05; β=0.10)

P_1	P_2	$\bar{p}$	2N
0.98	0.92	0.95	448
0.95	0.85	0.90	305
0.95	0.77	0.86	124
0.60	0.40	0.50	210
0.60	0.30	0.45	90
0.70	0.30	0.50	50

For example, for the risk factors with high probability, we require very large sample sizesto observe group differences. Suppose the overall probability is 0.95 and we wish to be able to observe a difference of 6% (one group would have p=0.98 and the other p=0.92). We would need a total sample size of 448 patients to detect this differencereliably. With an overall observed probability of 0.90, we would need 305 patients to see a difference of 10%.

For some of the lower awareness risk factors, we are able to observe differences (if they are in fact there) with lower sample sizes. The overall awareness of, for example, low fibre dietwas around 0.50. To observe a difference of 20% would require only 210 patients. To observe a difference of 30% would require just 90 patients.

If we decrease the power to 0.80, the sample sizes are reduced (Table 19) but this means that we are less likely to observe a significant difference if itis there (that is, the probability of correctly rejecting the null hypothesis has been reduced.)

Table 19 - Sample Size Calculations
(α=0.05; β=0.20)

P_1	P_2	$\bar{p}$	2N
0.98	0.92	0.95	324
0.95	0.85	0.90	220
0.95	0.77	0.86	124
0.60	0.40	0.50	152
0.60	0.30	0.45	66
0.70	0.30	0.50	36

5.5.2. How could we obtain a larger sample size?

As described previously, this study was designed as a census of the population on a specific day at the UOHI. To increase our sample size and hence the power of the dataset to observe relationships, it is clear from above that a greater sample size would be required.

One option would be to increase the population through a multi-centre trial. For example, the study could include all Eastern Ontario cardiac care patients admitted to hospital. Alternately, the definition of "patient" could be extended to include individuals in other types of institutions (e.g., long term care) or living at home. Criteria would have to be devised to ensure a common definition of individuals to be included (e.g., diagnosed with one of a set of diseases, taking one or more of a set of medications). Finally, the study could continue to be confined to the UOHI but include all admissions over a certain time period or until sufficient sample had been obtained. Clinical trials are often conducted this way, in that subjects are admitted to the study until a pre-defined sample size is obtained. In the case of this study, if we were able to repeat the study once per week and recruit 40 new patients each week, it would take 11 weeks to obtain the 448 patients

required as per the top line of Table 18. This is conservative as it ignores the additional data that would be required to model any time dependence.

The above discussion ignores the question of the cost/benefit ratio of increasing the sample size. Given the objectives of the research, it is not clear that the benefits of increased precision of estimates or discovery of new relationships in the data would be commensurate with the increased costs of data collection. This is a very real issue in health research at present, in particular, for studies funded by health care institutions and not, for example, by pharmaceutical companies.

6 Conclusions and Recommendations for Further Study

As stated in Chapter 1, Research Goal 1 was to understand the level of knowledge of cardiac risk factors in the UOHI patient population. While most patients were unable to identify many if any risk factors without prompting, when presented with risk factors in a multiple choice format, they were able to identify an average of 13.6 out of 16 risk factors. This is quite a high level of general knowledge, and is on par with the literature.

This study found differences in overall knowledge level depending on the risk factor of interest, from a high of 97% for high blood pressure to a low of 49% for low fibre diet. These results were published (Appendix 4) and disseminated within the UOHI.

The bulk of the thesis considered Research Goal 2: what subgroups of the patient population could benefit from additional educational initiatives, i.e., had lower than average knowledge levels? In general, the descriptive (and in general univariate) analysis suggests that younger (30's and 40's) males form the patient group most needing additional information, as well as individuals who did not complete high school.

The multidimensional analysis yielded additional insight into this second research goal. Logistic regression results for the low fibre diet risk factor identified education as a significant predictor. After removing an influential observation, education was also a significant predictor variable in logistic regression models for high salt diet and inactivity. Finally, smoking status was also significant in the model for inactivity with current smokers significantly less aware of the risks of inactivity than former and never smokers.

The correspondence analysis approach provided another level of information about relationships within the dataset. Once again, the relationship between knowledge of a low fibre diet and education was evident, and we noted in addition that gender was clearly not an issue for this risk factor. We also found that women were more aware of the risks of high salt diet and stress than men, and were also more likely than men to have never smoked. There was also a triad of smoking, lower education, and lack of awareness of the risk of a low fibre diet.

Finally, knowledge of the top three risk factors (smoking, high blood pressure, and hypertension) was apparently not related to either gender or education.

As noted in Chapter 5.2.4, logistic regression was hampered by small sample size and high probabilities for many risk factors. As discussed in Section 5.5, larger sample size would have improved our ability to find significant predictors in a similar (but larger) sample.

In Section 5.4, we discussed the thesis goal, comparing the two approaches to the analysis of this dataset. We summarize that discussion in Figure 22. In this study, the salient research goal was the identification of subgroups of the UOHI population that might benefit from additional educational intervention. For these purposes, the identification of relationships between low CRF knowledge and demographic variables was more important in and of itself than in quantifying these relationships and, given the relatively small sample size and sparseness of the data, the correspondence analysis produced more actionable information than the logistic regression approach.

Figure 22 - Comparison of Analytical Approaches

Criteria	Logistic Regression	Correspondence Analysis
Category of approach	Quantitative	Qualitative
Results obtained	Few significant relationships identified	More relationships visualized. May not be significant in traditional sense.
Ability to deal with sparse data	Limits models that can be estimated	No limitations
Ability to deal with small sample size	Reduces ability of data to demonstrate relationships	No impact
Usefulness of results	Can make quantitative statements about relationships Good for: Regulatory work Demand estimates Comparison with previous work	Directional only. Cannot quantify relationships. Good for: Exploratory work "Soft" questions
Ability to address stated research goals	Yes	Yes

There are a number of areas for further research, for both the research goals and thesis goals. This study collected a huge amount of data, of which only a small portion was included in the analysis for this thesis. For example, we have not touched on the knowledge of significant others, nor have we included in this investigation many of the variables collected in the chart audit (for example, Body Mass Index, medication profiles, diagnosis.)

With respect to the thesis goal, it would be very interesting to repeat this comparison with a larger dataset in which the logistic regression produces more significant variables. This would make comparison of the two approaches much more interesting and informative.

There are some other advanced techniques for dealing with the deficiencies in the data. A Bayesian approach to logistic regression using the significant other responses as prior information might improve the ability to identify significant predictors. More advanced techniques for estimating models with missing data are also available (for example, the EM algorithm[21]) as are robust techniques such as M-estimation for estimating models with cleaned data.

A more comprehensive study on the subject of risk factor knowledge could be conducted via a survey of the general population. This would permit much larger sample sizes and yield information about healthy people of all ages as well as those with cardiac problems. Because cardiac risk factors develop over an extended period of time, it is crucialto address the young and healthy with educational initiatives to achieve the goal of primary prevention.

Appendices

1. Survey instruments and Ethics Approval
2. Data file(s)
3. Programs
4. Journal article
5. Correspondence Analysis notation and process

Appendix 1 – Survey Instruments

This appendix contains the following documents:

1. Patient Survey
2. Significant Other Survey
3. Chart Audit form
4. Chart Auditors Guide
5. Patient Information Sheet and Consent Form
6. Ethics Approval

Patient's unique ID _____ Interviewer's initials _____

Cardiac Risk Factor Study – Patient Interview Script

Addressograph Impression

Hello. Thank you for agreeing to participate in this study. As was described to you, we are conducting a study to explore people's understanding about heart health in general and about their own health in particular. We want to find out how we can improve the quality of care that we provide. There are no right or wrong responses, so feel free to ask questions or change your answers.

1. **Please tell me why you are in the hospital.**

2. *(Ask only if they have not already told you their diagnosis in response to Question 1.)*
   What is your diagnosis?

3. **How long have you been in the hospital?**

   Number of Days _____ or Day/Date of Admission _____

4. **Do you know approximately how long you are expected to be in hospital at the Heart Institute? (If yes, how long?).**

   Number of Days _____ or Expected Day/Date of Discharge _____

5. **Who is your family physician?**

6. **We know that the risks for heart disease vary by ethnic background. Please tell me what your ethnic background is.**

7. **Out of the following categories, how would you classify your educational level?** (*read list*)

☐ Did not complete high school
☐ Completed high school
☐ Completed college
☐ Completed university
☐ Other (*please specify*) _____

8. **Where will you be going after you leave the Heart Institute?** (*do not read list*)

☐ Home
☐ Back to a regional hospital
☐ Home of friend or family member
☐ Don't know
☐ Other (*please specify*) _____

9. **As I mentioned previously, this study is looking at cardiac risk factors. Could you please tell me how you would define 'cardiac risk factor'?** (*Record verbatim response*)

Thank you. For this study, we will define a cardiac risk factor as something that makes it more likely that you will develop heart disease.

10. **Thinking now in general, what would you say the risk factors are for heart disease?**

11. **Where did you get your information about these general risk factors?**
(*Read list of sources one by one and for those used, ask satisfaction question.*)

Information Source	✓	**Were you satisfied with the information you received?**	
		Yes	**No**
Family doctor			
Heart specialist			
Nurse(s) at Heart Institute			
Nurse elsewhere – Specify:			
Friends or family			
The Internet			
Books			
Television, newspapers, magazines (*circle*)			
Other – Specify:			

12. **Can you tell me what your <u>personal</u> risk factors for heart disease are?**

13. **Where did you get your information about <u>your</u> risk factors?**
(Read list of sources one by one and for those used, ask satisfaction question.)

Information Source	✓	Were you satisfied with the information you received?	
		Yes	No
Family doctor			
Heart specialist			
Nurse(s) at Heart Institute			
Nurse elsewhere – Specify:			
Other – Specify:			

Only after the previous questions have been answered, proceed to Question 14.

14. **For each of the following conditions, please tell me if you think it is a risk factor for heart disease in general. You can answer "yes", "no", or "don't know".**

(For each condition, record first whether the patient thinks it is a RF for HD. Then go back to the items that they identified as cardiac risk factors and ask them whether it is a personal risk factor, and record any additional information.)

Condition	Risk Factor for Heart Disease? Yes	No	DK	Personal Risk Factor?	If a personal risk factor, ask for additional information
Being over age 35 for men, age 45 for women					
History of broken leg					
Family history of heart disease					Complete family history chart on next page. ❶
High blood cholesterol					
Having had a general anaesthetic					
A diet high in salt					
High blood sugar					
A diet low in fibre					
Having had previous surgery					
Being overweight					
Gout					
Being a smoker					Complete smoking history on next page. ❷
Over age 45 years for men, age 55 for women					
A diet high in fat					
Having high blood pressure					
Having hayfever					
Lack of physical activity					Regardless of the answer, fill in information on physical activity. ❸

Condition	Risk Factor for Heart Disease?			Personal Risk Factor?	If a personal risk factor, ask for additional information
	Yes	No	DK		
Having had previous heart problems					Whether they consider this a risk factor or not, find out if they have had previous heart problems. ❹
Having a history of cancer					
Being exposed to second hand smoke					**How are you exposed to second hand smoke?**
Having had chicken pox					
Stress					
Having a history of rheumatic fever					
Being a former smoker (quit within past 2 years)					Clarify if individual is a former smoker (even if quit >2 years ago) and complete smoking history. ❷
For women only: Being post-menopausal					

❶ **(Family) History of Heart Disease**

Who in your family has had heart disease? (*Use reverse if required*)

Relationship (Mother, brother…)	What was their diagnosis?	At what age did they start to get symptoms?

❷ **Smoking History**

For smokers: **How long have you smoked:**	*For smokers and former smokers:* **What type? (cigarettes, cigars, pipe, other)**
For former smokers: **How long did you smoke?** **How long ago did you quit?**	**How much do you smoke per day?**

❸ **Physical Activity**

What type of physical activity do you do? How often and for how long?

Could you tell me what type of physical activity a person needs to engage in in order to reduce their risk of heart disease?

❹ **Patient's History of Heart Disease**

Have you had previous heart problems? Yes ___ No ____

(If the answer is 'yes', fill in the table below.) (Use reverse if required)

What was their diagnosis?	What were their symptoms?	At what age did they start to get symptoms?	What treatment did they receive?

15. **Finally, there is a lot of discussion in the scientific literature on alcohol and the amounts that affect the heart.**

Do you consume alcohol?

☐ No

☐ Yes → **How many drinks per day?** _____ or

How many drinks per week? _____

Of what type?

Thank you for your participation. Do you have any questions regarding the survey for me?

(If yes, summarize what was asked and how you responded. Use the back of this page if you need to.)

☐ Patient requested further information about cardiac risk factors. Yes _____ No _____

☐ Patient would like a summary of the research results mailed to them. Yes _____ No _____

Patient's unique ID _____ Interviewer's initials _____

Cardiac Risk Factor Study – Significant Other Interview Script

Definition of Significant Other
For the purposes of this study, the significant other is the person most likely to be involved in the care and recovery of the patient once they leave the Heart Institute.

Hello. Thank you for agreeing to participate in this short quality study.

1. **Please tell me what your relationship is to the patient.**

2. **How long have you known them?**

3. **How often do you expect to be involved in the care and recovery of [the patient]?**
   ☐ Daily (lives with the patient)
   ☐ Daily (will drop in to see them)
   ☐ Once a week
   ☐ Other (*please specify*) _____

4. **What do you expect your involvement will include?**

5. **Please tell me why [the patient] is in the hospital.**

6. *(Ask only if they have not already told you the patient's diagnosis in response to Question 5.)*
   What is [the patient's] diagnosis?

7. **How long has [the patient] been in the Heart Institute?**

8. **Approximately how long are they expected to be in hospital at the Heart Institute?**

   _____ days or

   ☐ Don't know

9. **Do you know where [the patient] will be going after they leave the Heart Institute (e.g. home, back to a regional hospital, etc.)?** (*do not read list*)

☐ Home
☐ Back to a regional hospital
☐ Home of friend or family member
☐ Don't know
☐ Other (*please specify*) _____

10. **As I mentioned previously, this study is looking at cardiac risk factors. Could you please tell me how you would define 'cardiac risk factor'?** (*Record verbatim response*)

Thank you. For this study, we will define a cardiac risk factor as something that makes it more likely that a person will develop heart disease.

11. **What do you think, in general, the risk factors are for heart disease?**

12. **Where did you get your information about these general risk factors?**
(*Read list of sources one by one and for those used, ask satisfaction question.*)

Information Source	✓	Were you satisfied with the information you received?	
		Yes	No
Family doctor			
Heart specialist			
Nurse(s) at Heart Institute			
Nurse elsewhere – Specify:			
Friends or family			
The Internet			
Books			
Television, newspapers, magazines (circle)			
Other – Specify:			

13. **Can you tell me what [the patient's] particular risk factors for heart disease are?**

14. **Where did you get your information regarding [the patient's] specific cardiac risk factors?**

Information Source	✓	Were you satisfied with the information you received?	
		Yes	No
Family doctor			
Heart specialist			
Nurse(s) at Heart Institute			
Nurse elsewhere – Specify:			
Other – Specify:			

Only after the previous questions have been answered, proceed to Question 15.

15. **For each of the following conditions, please tell me if you think it is a risk factor for heart disease in general. You can answer "yes", "no", or "don't know".**

(For each condition, record first whether the SO thinks it is a RF for HD. Then go back to the items that they identified as cardiac risk factors and ask them whether it is a risk factor the patient has, and record any additional information.)

Condition	Risk Factor for Heart Disease?			Personal Risk Factor?	If a personal risk factor, ask for additional information
	Yes	No	DK		
Being over age 35 for men, age 45 for women					
History of broken leg					
Family history of heart disease					Complete family history chart on next page. ❶
High blood cholesterol					
Having had a general anaesthetic					
A diet high in salt					
High blood sugar					
A diet low in fibre					
Having had previous surgery					
Being overweight					
Gout					
Being a smoker					Complete smoking history on next page. ❷
Over age 45 years for men, age 55 for women					
A diet high in fat					
Having high blood pressure					
Having hayfever					
Lack of physical activity					Regardless of the answer, fill in information on physical activity. ❸

Condition	Risk Factor for Heart Disease?			Personal Risk Factor?	If a personal risk factor, ask for additional information
	Yes	No	DK		
Having had previous heart problems					Whether they consider this a risk factor or not, find out if the patient had previous heart problems. ❹
Having a history of cancer					
Being exposed to second hand smoke					**How were they exposed to second hand smoke?**
Having had chicken pox					
Stress					
Having a history of rheumatic fever					
Being a former smoker (quit within past 2 years)					Clarify if the patient is a former smoker (even if quit >2 years ago) and complete smoking history. ❷
For women only: Being post-menopausal					

❶ **(Family) History of Heart Disease**

Who in your family has had heart disease? (*Use reverse if required*)

Relationship (Mother, brother…)	What was their diagnosis?	At what age did they start to get symptoms?

❷ **Smoking History**

For smokers: **How long have they smoked:**	*For smokers and former smokers:* **What type? (cigarettes, cigars, pipe, other)**
For former smokers: **How long did they smoke?** **How long ago did they quit?**	**How much do they smoke per day?**

❸ **Physical Activity**

What type of physical activity does the patient do? How often and for how long?

Could you tell me what type of physical activity a person needs to engage in order to reduce their risk of heart disease?

❹ Patient's History of Heart Disease

Has the patient had previous heart problems? Yes ___ No ____

(If the answer is 'yes', fill in the table below.) (Use reverse if required)

What was their diagnosis?	What were their symptoms?	At what age did they start to get symptoms?	What treatment did they receive?

16. **Finally, there is a lot of discussion in the scientific literature on alcohol and the amounts that affect the heart.**

Does the patient consume alcohol?

☐ No

☐ Yes → **How many drinks per day?** _____ or

How many drinks per week? _____

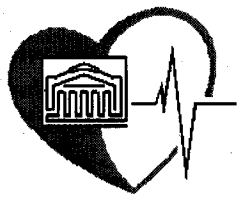
Of what type?

Thank you for your participation. Do you have any questions regarding the survey for me?

(If yes, summarize what was asked and how you responded. Use the back of this page if you need to.)

☐ SO requested further information about cardiac risk factors. Yes _____ No _____

☐ SO would like a summary of the research results mailed to them. Yes _____ No _____



OTTAWA HEART INSTITUTE

Cardiac Risk Factor Audit Form

Date of Chart Audit _____ Initials of Auditor _____

From Hospital Admission Report

Surname, Name _____ Unique # _____

Admission Date _____ Time _____

DOB (yy/mm/dd)_____ Sex _____

Preferred Language _____

Address _____ Phone _____

_____

Name of significant other or caregiver _____ _____

Relationship _____ Phone _____

Family Physician _____ Phone _____

Heart Inst Physician _____

From Medical History or Registry Form

Admission diagnosis _____

NYHA Functional Classification _____ **CCS Angina Class** _____

From Clinical Flow Sheets

Vital signs upon admission: B.P. _____ HR _____ Respirations _____

- ☐ **Elevated BP**
- ☐ **Hypertension**

Nursing History or Medical History

Height _____ **Weight** _____ **BMI** _____

Alcohol consumer

- ☐ No
- ☐ Yes ⇨ Drinks per day _____ Type _____ or

  Drinks per week _____ Type _____

Family history of heart disease

- ☐ No
- ☐ Yes

Family member 1 _____ Diagnosis
Age of onset of cardiac symptoms _____

Family member 2 _____ Diagnosis
Age of onset of cardiac symptoms _____

If there are others, use reverse.

Smoking

- ☐ Never
- ☐ Former
- ☐ Current

Number smoked per day (now or when quit)_____
Length of time as smoker (yy/mm) _____
Number smoked per day _____ **Type** _____
Quit date (yy/mm) *(Former Smokers)* _____

Other cardiac risk factors

- ☐ **Poor cardiac diet:** Diet high in fat
- ☐ **Poor cardiac diet:** Diet low in fibre
- ☐ **Diabetes**
- ☐ **Elevated B.P**
- ☐ **Hypertension**
- ☐ **Sedentary**
- ☐ **Previous cardiac event** ⇨ What and when _____
- ☐ **Stress** ⇨ What type? _____
- ☐ **Post-menopausal (for women)**

Other comments

From Lab Work

Blood glucose _____

(Complete sections below…. within 24 hour preferred)

Check one:

- ☐ Fasting
- ☐ Non-fasting

Check one:

- ☐ Within 24 hr
- ☐ >24 hr

Cholesterol

(Complete sections below…. within 24 hour preferred)

Check one:

- ☐ Fasting
- ☐ Non-fasting

Check one:

- ☐ Within 24 hr
- ☐ >24 hr

Total		mmol/L
HDL		mmol/L
Triglycerides		mmol/L
LDL		mmol/L
TC/HDL Ratio		

From Nursing History, Medical History or MAR

Place ✓in appropriate columns for medication. List other meds on at end of table (next page).

Drug Category	*Drug Name*	*On Admission*	*In Hospital*	*At discharge*
Beta Blockers	Metropolol			
	Atenolol			
	Nadolol (Corgard)			
Other (specify)				
ACE Inhibitors	Ramapril			
	Vasotec			
Other (specify)				
Ca Channel Blockers	Diltiazem			
	Amlodipine			
Other (specify)				
Angiotensin ll receptor blockers	Losarten (Cozaar)			
	Valsartan (Diovan)			
	Irbesartan (Avapro)			
Other (specify)				
Other Cardiac and Anti-hypertensives	Hydralazine			
Other (specify)				
Nitrates	Nitrodur			
	Isordil			
Other (specify)				
Lipid Metabolism Regulators	Simvastatin (Zocor)			
	Atorvastatin (Lipitor)			
	Fenofibrate (Lipidil)			
	Pravastatin (Pravachol)			
Other (specify)				
Diuretics	Lasix			
	Hydrochlorothiazide			
Other (specify)				
Anti-platelet	ASA			
	Clopidogrel (Plavix)			
	Eptifibatide (Intregrelin)			
Other (specify)				

Drug Category	*Drug Name*	*On Admission*	*In Hospital*	*At discharge*
Anti-coagulants	Coumadin			
	Heparin			
	Enoxaparin			
	Fragmin			
Other (specify)				
Hormone Replacement	(Women only)			
	Morphine			
	Scapolamine			
	Ativan			

Chart Auditor's Guide

Alcohol consumption

The literature varies on this point. Get as many details as possible.
One drink is 1½ oz of hard liquor or 5 oz wine or 12 oz beer.

1. **Family history of heart disease**
   Includes all first-degree relatives (parents, siblings, children) and second-degree

2. **Smoking**
   Be as specific as possible.

3. **Poor cardiac diet**

   Elevated intake of dietary fats and/or diet low in fibre.
   (Most likely place to find this is in the nutritionist's notes if there was a nutrition consult).
   Example of a high fat diet:
   >30% total calories from fat
   >10% total calories from saturated fat
   >300 mg. cholesterol per day
   Example of a low fibre diet: Few or no fruits or vegetables, bran or other high fibre foods.

4. **Diabetes**

   Identified as diabetic either in the physician's history, the nurses' notes or the Kardex.

5. **Elevated blood pressure**
   Have 3 consecutive elevated B.P.s (systolic > 140and/or diastolic > 90) within

6. **Hypertension**

Identified as hypertensive either in the physician's history, the nurses' notes or the Kardex.

7. **Sedentary**
   Does not participate in some form of aerobic activity for at least 20 minutes at least 3 times a week. Walking only counts if they walk fast enough to increase their heart rate.

8. **Previous cardiac event**

9. **Stress**
   Identified on the chart or kardex as a problem for this patient or as a risk factor.
   Life events that could reasonable be considered very stressful but not identified as

10. Post-menopausal

For the purposes of this audit, this only includes conventional HRT, not homeopathic drugs.

UNIVERSITY OF OTTAWA HEART INSTITUTE / INSTITUT DE CARDIOLOGIE DE L'UNIVERSITÉ D'OTTAWA
40 rue Ruskin Street, Ottawa, Ontario K1Y 4W7 www.ottawaheart.ca

PATIENT INFORMATION SHEET AND CONSENT

Cardiac Risk Factor Point Prevalence Study

INVESTIGATOR: **Dr. K. L. Momtahan, R.N., Ph.D.**
CO-INVESTIGATORS: **Sharon Ann Kearns RN, BScN**
Nancy Lauzon RN, BScN
Heather Sherrard , R.N., MHA
Janet Berkman, Masters Student in Epidemiology
Judith Sellick, RN, BScN
STUDY COORDINATOR: **Dr. K. L. Momtahan, R.N., Ph.D.**

Please read this Patient Information Sheet and Consent Form carefully and ask as many questions as you like before deciding whether to participate.

INTRODUCTION

We are conducting a one-day information-gathering exercise to investigate the level of understanding that patients have regarding cardiac risk factors in general,and their own risk factors in particular, by conducting a short interview with them. We would also like to interview the person most likely to be involved in your recovery once you go home. This information will help us decide if we need to change how we approach education for patients with regards to cardiac risk factors.

PROCEDURE

We are asking all patients who are able to participate in this study on June 5. The study itself will run from 7:30 a.m. to 7:30 p.m.. Your participation in the study should take no more than about a half hour.

There are three parts to the study:

1. A short interview that you will have with one of the volunteer nurses.
2. An audit of your chart that another volunteer nurse will carry out.
3. A short interview with the person in your life who will have the most involvement with your recovery once you leave the Heart Institute. This might be your spouse, a son or daughter or a close friend, for example. If this person is not at the Heart Institute on June 5, we can do a telephone interview with them as long as we are able to do so by June 12.

All the nurses who have volunteered to conduct this short study are employees of the University of Ottawa Heart Institute.

RISKS and DISCOMFORTS of PARTICIPATION

There are no additional risks or discomforts involved in this study. If you are too tired to be interviewed at the time that you are approached, but you would like to participate in the study, we can arrange for the nurse to come back later in the day to interview you.

BENEFITS OF PARTICIPATION

There are no direct benefits of this study to you. However, we are hoping that the information we get will improve how we approach the education of patients with regards to cardiac risk factors.

COMPENSATION/RENUMERATION

There is no monetary compensation to being in the study. By participating in the study and signing the consent form, you are not waiving your legal rights which may be available to you.

CONFIDENTIALITY

All records will be kept confidential. Your health records and the interviews will be reviewed by the Investigators and may be reviewed by the representatives of the Heart Institute Research Ethics Board under the supervision of the Investigator or her staff. You will not be identified in any publications by name or initials. All interview sheets will have your number only on them and they will be kept in a locked cupboard at the Heart Institute.

TERMINATION OF THE STUDY

Your participation is voluntary, and you may elect to refuse to participate or to discontinue participation at any time.

ETHICS

This study has been approved by the Human Research Ethics Board of the University of Ottawa Heart Institute. The Board considers the ethical aspects of all Heart Institute research projects using human subjects. If you wish, you may talk to the Chair of the Human Research Ethics Board through the Secretariat at (613) 761-4417.

PARTICIPATION **Participation in research is completely voluntary.**

You are free to choose whether to participate in this study or not. If you choose to participate, you may choose to withdraw your consent at any time. This will not prejudice your present and future care at this hospital. You are also free to refuse to answer anyquestions that you may be asked because of your participation in this study.

CONSENT TO PARTICIPATE IN RESEARCH

I understand that I am being asked to participate in a research study on cardiacrisk factors. By giving my consent, I am authorizing the University of Ottawa Heart Institute personnel to screen my medical record for the purposes of the cardiac risk factor point prevalence study. I am also agreeing to be interviewed for the purposes of this study.

I have read and understood this Patient Information Sheet and Consent Form. All of my questions at this time have been answered to my satisfaction. If I or any of my family members have any further questions about this study, we may contact the Principal Investigator Dr. Kathryn Momtahan, R.N., Ph.D. at (613) 798-5555, x 13575.

I will receive a signed copy of this Consent Form and the attached Patient Information Sheet.

I voluntarily agree to participate in this study.

Patient Name _____

Patient Signature _____

Date _____

Witness Signature _____

Signature of person obtaining Consent _____
Investigator / Co-Investigator's Signature _____

Name of family member/friend participating in the study
_____

Signature of family member/friend participating in the study
_____

Date _____

Witness Signature _____

Signature of person obtaining Consent _____
Investigator / Co-Investigator's Signature _____

UNIVERSITY OF OTTAWA HEART INSTITUTE / INSTITUT DE CARDIOLOGIE DE L'UNIVERSITÉ D'OTTAWA
40, rue Ruskin Street, Ottawa, Ontario, Canada K1Y 4W7 www.ottawaheart.ca

April 18, 2002

Dr. Kathy Momtahan
Clinical Services
University of Ottawa Heart Institute

UOHI - 02 - 121 Cardiac Risk Factor Prevalence Study

Dear Dr. Momtahan:

Subsequent to the HREB expedited review, all outstanding issues have been addressed and the required revisions have been made. I am pleased to advise you that this project now has received **FULL APPROVAL**, valid for a period of one year ending on April 18, 2003.

Should you wish for project approval to be extended, a Continuing Review Form should be completed and sent to the HREB at least one month prior to the expiration date. Should the study terminate during the year, please complete and forward a Termination Report to the HREB.

As a reminder, the Tri-Council Policy requires a continuing involvement of research ethics committees in studies over the course of their execution. You must maintain, as part of your records, the original signed Consent form. As well, you must inform the HREB of adverse events encountered during the study, here or elsewhere, or of significant new information which becomes available after the HREB review, which may impinge on the ethics of continuing the study. The HREB will review the new information to determine if the protocol should be modified, discontinued, or continue as originally approved.

Please note that the HREB requires all Patient Information Sheets & Consent Forms to be available in both official languages. Please forward a copy of the French version of the Consent to the HREB as soon as it is translated for validation.

Yours sincerely,

Dr. J. George Fodor
Chair,
Human Research Ethics Board

Encl.

Appendix 2 – Data File

This appendix contains raw data used in the analysis. This is a subset of a much larger dataset, and the programs that follow in Appendix 3 must be modified to accommodate this reduced dataset. Variable definitions and coding appear at the end of the dataset.

OBS	GENDER	AGE	LANG	EDUCAT7	Q14_1GE	Q14_2GE	Q14_3GE	Q14_4GE	Q14_5GE	Q14_6GE	Q14_7GE	Q14_8GE	Q14_9GE	Q14_10GE	Q14_11GE	Q14_12GE	Q14_13GE	Q14_14GE	Q14_15GE	Q14_16GE	Q14_17GE	Q14_18GE	Q14_19GE	Q14_20GE	Q14_21GE	Q14_22GE
1	f	74	e	hs	y	n	y	y	n	y	y	y	y	y	n	y	y	y	y	dk	y	y	n	y	n	y
2	m	77	e	col	dk	n	y	y	n	y	n	n	n	y	y	y	y	y	y	n	y	y	n	y	n	y
3	f	39	e	hs	y	y	y	y	n	y	y	y	n	y	dk	y	y	y	y	n	y	y	n	y	n	y
4	m	67	e	nohs	y	n	y	y	n	y	y	n	n	y	n	y	y	y	y	n	n	y	y	y	n	y
5	f	34	e	nohs	dk	y	y	dk	y	y	y	dk	dk	y	dk	y	y	y	dk	dk	y	dk	y	y	dk	y
6	m	51	e	col	n	n	y	y	dk	y	y	n	n	y	dk	y	y	y	y	dk	y	y	dk	dk	dk	y
7	m	62	e	nohs	n	n	y	y	n	y	y	n	n	y	n	y	y	y	y	n	y	y	dk	y	n	y
8	m	86	e	hs	n	n	y	y	n	y	y	n	n	y	n	y	y	y	y	dk	y	y	dk	y	n	y
9	m	48	e	col	y	n	y	y	n	y	y	y	n	y	dk	y	y	y	y	n	y	y	n	y	n	y
10	m	64	e	hs	y	n	y	y	dk	y	y	dk	dk	y	n	y	y	y	y	dk	y	y	dk	y	n	y
11	m	59	e	nohs	n	n	y	y	n	n	y	n	n	y	n	y	y	y	y	dk	y	y	y	n	n	n
12	f	80	e	hs	y	y	y	y	dk	y	y	n	n	y	y	y	y	y	y	y	dk	y	y	y	dk	y
13	f	65	e	nohs	y	n	y	y	dk	y	dk	dk	dk	y	dk	y	y	y	y	dk	dk	y	dk	y	dk	y
14	m	56	e	nohs	n	n	y	y	y	y	y	y	y	y	dk	y	y	y	y	dk	y	y	dk	y	dk	y
15	m	76	e	nohs	dk	n	y	y	n	n	y	y	n	y	dk	y	n	y	y	dk	y	y	dk	y	dk	y
16	m	59	e	col	dk	n	y	y	n	dk	y	y	n	y	dk	y	y	y	y	n	y	y	n	y	dk	dk
17	f	58	e	uni	y	n	y	y	y	y	y	y	y	y	n	y	y	y	y	n	y	y	n	y	n	y
18	m	63	e	nohs	n	n	y	y	dk	y	y	dk	dk	y	dk	y	y	y	y	dk	y	y	dk	y	dk	y
19	f	72	e	col	n	n	y	y	dk	y	y	y	n	y	dk	y	y	y	y	dk	y	y	dk	y	dk	y
20	m	81	e	hs	dk	n	dk	dk	dk	dk	dk	dk	dk	y	dk	dk	dk	dk	dk	dk	dk	dk	dk	y	dk	y
21	m	78	e	uni	y	n	y	y	n	y	y	y	n	y	n	y	y	y	y	dk	y	y	dk	y	dk	y
22	m	80	e	hs	y	y	y	y	n	y	y	y	y	y	dk	y	y	y	y	n	y	y	dk	y	dk	y
23	m	81	e	nohs	dk	n	y	y	y	dk	y	dk	y	y	y	y	dk	dk	y	y	y	y	y	y	dk	y
24	m	70	e	nohs	n	n	y	y	n	n	dk	dk	dk	y	dk	y	n	y	y	dk	y	y	n	y	dk	y
25	f	55	e	uni	y	n	y	y	n	y	y	y	y	y	n	y	y	y	y	n	y	y	dk	y	n	y
26	f	56	e	hs	n	n	y	y	n	y	y	y	dk	y	n	y	n	y	y	n	y	y	n	y	n	y
27	m	79	e	nohs	n	n	y	y	n	y	y	dk	y	y	y	y	y	y	y	dk	y	y	n	y	dk	y
28	f	70	e	nohs	n	n	y	y	y	y	y	dk	y	y	n	y	y	y	y	n	y	y	n	y	n	y
29	m	67	f	uni	y	n	y	y	n	y	n	n	y	y	dk	n	y	y	y	y	y	y	n	n	dk	y
30	m	67	e	uni	y	n	y	y	n	y	y	n	y	y	n	y	y	y	y	dk	y	y	y	y	n	y
31	m	60	e	hs	n	dk	y	y	n	y	y	dk	y	y	n	n	y	y	y	dk	y	y	n	y	n	y
32	m	62	e	uni	y	n	y	y	dk	y	y	y	y	y	y	y	y	y	y	n	y	y	n	n	n	y
33	m	79	e	hs	y	n	y	y	n	y	y	y	n	y	y	y	y	y	y	n	y	y	n	y	y	y
34	m	53	e	nohs	y	n	y	y	dk	y	y	y	dk	y	n	y	y	y	y	dk	y	y	y	y	dk	y
35	f	71	e	hs	y	n	y	y	dk	y	y	y	n	y	y	y	y	n	y	dk	y	dk	y	y	n	y
36	m	59	e	col	y	n	y	y	n	y	y	y	n	y	n	y	y	y	y	n	y	y	y	y	dk	y
37	m	83	e	nohs	y	dk	y	y	n	y	y	n	n	y	dk	y	y	y	y	dk	n	y	y	n	n	y
38	f	70	e	nohs	y	n	y	y	n	y	y	n	n	y	dk	y	y	y	y	n	y	y	n	y	dk	n
39	f	79	f	nohs	n	n	y	y	n	y	y	y	n	dk	dk	y	y	y	y	dk	y	y	n	y	n	y
40	m	58	e	nohs	n	n	y	y	n	n	n	n	n	n	dk	y	y	y	y	n	n	y	n	y	n	y
41	f	48	e	hs	n	n	y	y	n	y	y	y	n	y	n	y	n	y	y	n	y	y	n	y	n	y
42	m	82	e	nohs	y	dk	y	dk	dk	y	y	dk	dk	y	dk	y	y	y	y	dk	y	y	dk	y	y	y
43	f	74	e	nohs	n	n	y	y	y	y	y	n	n	y	y	y	n	y	y	n	y	y	n	y	n	y
44	f	56	e	col	y	n	y	y	dk	y	y	y	n	y	y	y	y	y	y	n	y	y	n	y	n	y

OBS	GENDER	AGE	LANG	EDUCAT7	Q14_1GE	Q14_2GE	Q14_3GE	Q14_4GE	Q14_5GE	Q14_6GE	Q14_7GE	Q14_8GE	Q14_9GE	Q14_10GE	Q14_11GE	Q14_12GE	Q14_13GE	Q14_14GE	Q14_15GE	Q14_16GE	Q14_17GE	Q14_18GE	Q14_19GE	Q14_20GE	Q14_21GE	Q14_22GE
45	m	59	e	nohs	y	n	n	y	n	y	n	y	n	y	n	y	y	y	y	n	y	y	y	y	dk	y
46	m	47	e	nohs	y	n	y	y	dk	n	y	dk	n	n	n	y	y	n	y	y	n	y	n	n	n	n
47	f	69	e	hs	y	n	y	y	dk	y	dk	dk	n	y	n	y	y	y	y	n	y	y	n	y	n	y
48	m	55	e	hs	y	n	y	y	dk	y	y	n	n	y	n	y	y	y	y	n	y	y	y	y	n	y
49	f	48	e	nohs	n	dk	y	y	dk	y	y	n	dk	y	dk	y	dk	y	y	dk	y	y	dk	dk	dk	y
50	f	63	e	nohs	y	n	y	y	n	y	y	y	n	y	n	y	y	y	y	dk	y	y	y	y	dk	y
51	m	53	e	col	y	y	y	y	n	y	y	n	y	y	n	y	y	y	y	n	y	y	n	y	n	y
52	f	80	e	nohs	n	n	y	y	n	y	y	dk	n	y	dk	y	n	y	y	dk	y	y	dk	y	n	y
53	m	31	e	nohs	y	n	y	y	dk	dk	n	y	dk	y	y	y	n		y	n	y	y	n	y	n	y
54	m	82	f	uni	dk	n	y	y	n	y	y	y	n	y	dk	y	y	y	y	dk	y	y	y	y	n	y
55	f	42	e	col	dk	n	y	y	dk	y	y	dk	n	y	y	y	y	y	y	dk	y	y	dk	y	dk	y
56	m	60	e	uni	n	n	y	y	n	y	y	y	n	y	n	y	y	y	y	n	y	y	n	y	n	y
57	m	85	e	uni	n	n	y	y	n	n	dk	n	n	y	y	y	n	y	y	n	n	y	dk	y	y	n
58	f	53	e	col	n	n	y	y	n	y	y	y	n	y	n	y	y	y	y	n	y	y	n	y	n	y
59	m	52	e	nohs	n	n	y	y	n	y	y	y	n	y	dk	y	n	y	y	n	n	y	n	y	n	y
60	m	57	e	col	y	n	y	y	n	y	y	y	n	y	y	y	y	y	y	n	y	y	n	y	dk	y
61	m	73	e	nohs	y	n	y	y	n	y	dk	y	n	y	dk	y	y	y	y	dk	n	dk	n	y	n	y
62	f	68	e	hs	n	n	y	y	dk	y	y	y	dk	y	y	y	n	y	y	dk	y	y	dk	y	dk	y
63	f	56	e	nohs	y	n	y	y	n	y	y	n	n	y	y	y	y	y	y	n	y	y	n	y	n	y
64	m	69	e	nohs	n	n	y	y	dk	y	y	dk	dk	y	dk	y	n	y	y	dk	y	dk	n	dk	dk	dk
65	m	68	e	uni	y	n	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y	dk	y
66	f	76	f	uni	n	n	n	y	dk	y	y	y	dk	y	y	y	y	y	y	n	y	y	dk	y	dk	y
67	m	52	e	hs	n	n	y	y	n	y	y	y	y	y	dk	y	y	y	y	n	y	y	n	y	n	y
68	m	42	e	nohs	y	dk	y	y	dk	y	y	n	dk	y	dk	y	y	y	y	n	y	y	dk	y	dk	y
69	f	55	e	col	y	n	y	y	n	y	y	n	n	y	n	y	y	y	y	n	y	y	n	y	n	y
70	f	59	e	hs	y	n	y	y	n	y	y	y	n	y	dk	y	y	y	y	y	y	y	dk	y	n	y
71	m	50	e	col	n	n	y	y	dk	y	dk	y	y	y	n	y	y	y	y	n	y	y	dk	y	y	y

OBS	Q14_23GE	Q14_24GE	Q14_25GE	SMOKSTA	PREVPROB	whg1	whg2	whg3	whg4	whg5	whg6	whg7	whg8	whg9	whg10	whg11	whg12	whg13	whg14	whg15	whg16	whg17	whg18
1	y	y	n	former	n	1						1											
2	n	y		former	y						1	1											
3	y	y	y	never	y	1					1	1		1					1				
4	n	y	y	never	y							1											
5	dk	dk	dk	former	n																		
6	dk	dk	dk	current	n	1		1				1		1		1							
7	n	y		former	y							1											
8	y	y	y	never	n	1						1				1							
9	y	y	y	former	y		1		1			1			1				1				
10	dk	dk		former	y	1			1	1	1	1				1							
11	n	y		current	y			1			1	1		1								1	
12	y	y	dk	never	n				1			1											
13	dk	y	dk	current	n										1								
14	dk	y	dk	former	y	1						1				1							
15	y	y		former	n	1	1												1				
16	dk	y	n/a	former	y	1						1							1				
17	n	y	y	former	n		1		1		1	1			1				1				
18	y	y	dk	never	y				1			1				1							
19	y	y	dk	never	y							1							1				
20	dk	dk	dk	former	n																		
21	dk	y	y	former	n				1			1		1					1				
22	y	y	dk	former	y									1					1				
23	y	y	y	former	y																		
24	y	y	y	current	y	1																	
25	n	n	y	former	n	1						1				1			1				
26	y	y	n	former	y						1	1											
27	y	y		former	y		1	1				1										1	
28	n	y	y	former	y							1		1									
29	dk	n	dk	former	y	1																	
30	y	y	y	never	n		1					1			1				1				
31	y	n	n	former	n						1	1				1							
32	n	n	y	never	y							1		1		1			1				
33	y	y	y	former	n	1													1				
34	y	y	y	former	y														1				
35	n	y		former	y							1											
36	y	y	y	current	n		1				1	1							1				
37	dk	n	dk	former	n							1											
38	dk	y	dk	never	y									1		1							
39	y	y	n	former	n																		
40	n	n	n	current	y							1				1							
41	n	n	y	current	y		1					1											
42	y	y	y	never	n				1		1	1											
43	y	y	y	former	y	1																	
44	y	y	y	current	n							1		1		1							

OBS	Q14_23GE	Q14_24GE	Q14_25GE	SMOKSTA	PREVPROB	whg1	whg2	whg3	whg4	whg5	whg6	whg7	whg8	whg9	whg10	whg11	whg12	whg13	whg14	whg15	whg16	whg17	whg18
45	y	y	dk	former	n							1		1									
46	n	y	dk	current	y							1											
47	y	n	y	former	y						1					1							
48	y	y		former	n	1					1	1		1		1							
49	dk	dk	y	never																			
50	dk	y	y	former	n						1	1											
51		y	y	former	n	1	1					1											
52	y	n	n	former	n																		
53	n	y	y	former	n							1		1									
54	n	y	dk	never	y		1									1			1				
55	y	y		never	n		1		1		1				1	1							
56	dk	y	dk	former	y				1			1		1									
57	dk	y	n	former	y	1																	
58	y	y	y	former	n							1				1							
59	y	y	n	never	n							1		1					1				
60	n	y	y	former	y		1		1			1							1				
61	y	y		current	n							1											
62	y	y	y	former	n	1																	
63	y	y	dk	former	n	1					1	1		1									
64	dk	y		former	y							1											
65	y	y	dk	former	y	1						1				1							
66	y	y	y	never	y											1							
67	y	y	y	never	n	1						1				1			1				
68	dk	y	dk	current	n						1	1											
69	y	y	y	current	y		1					1				1							
70	y	y	y	former	y	1						1		1		1							
71	y	y	y	former	n							1							1				

OBS	whg19	whg20	whg21	whg22	whg23	whg24	whg25	whg26	whg27	whg28	whg29	whg30	whg31	whg32	whg33	whg34	whg35	whgdk
1																		
2																		
3		1																
4		1																
5																		1
6																		
7			1															
8		1							1									
9																		
10		1																
11																		
12		1																
13																		
14		1							1									
15				1														
16																		
17																		
18								1	1									
19		1							1									
20																		1
21																		
22				1														
23																1		
24																		
25									1					1				
26		1																
27				1														
28																		
29																		
30																		
31		1						1										
32		1																
33									1									
34									1	1				1				
35																		
36																		
37																		
38																		
39				1														
40									1									
41																		
42		1																
43																		
44									1					1				

OBS	whg19	whg20	whg21	whg22	whg23	whg24	whg25	whg26	whg27	whg28	whg29	whg30	whg31	whg32	whg33	whg34	whg35	whgdk
45																		
46																		
47									1									
48																		
49																		1
50									1									
51									1									
52																		1
53		1																
54				1					1									
55																		
56									1									
57																		
58									1									
59																		
60																		
61																		
62									1					1				
63			1															
64																		
65		1							1									
66																		
67									1									
68									1									
69									1					1				
70																		
71		1							1					1				

Data codes

Obs: Observation number

Gender: f(emale), m(ale)

Age: in years

Language: e(nglish), f(rench)

Educat7: nohs – did not complete high school
hs – completed high school
col – completed college
univ – completed university

smoksta: Smoking Status

never – never smoked
former –former smoker
current – current smoker

prevprob: "Have you had previous heart problems?" y(es), n(o)

Q14_xxGE: These refer to the responses to Question 14 on the patient interview : "is the following condition a risk factor for heart disease? " Possible responses: y(es), n(o), d(on't)k(now)

The conditions are:

1. Being over age 35 for women, age 45 for men (i.e., Variable Q14_1GE)
2. History of broken leg
3. Family history of heart disease
4. High blood cholesterol
5. Having had a general anaesthetic
6. A diet high in salt
7. High blood sugar
8. A diet low in fibre
9. Having had previous surgery
10. Being overweight
11. Gout
12. Being a smoker
13. Over age 45 for women, 55 for men
14. A diet high in fat
15. Having high blood pressure
16. Having hayfever
17. Lack of physical activity
18. Having had previous heart problems
19. Having a history of cancer
20. Being exposed to second hand smoke
21. Having had chicken pox
22. Stress
23. Having a history of rheumatic fever
24. Being a former smoker (quit within past 2 years)
25. (For women only) Being post-menopausal.

whg**xx**: This variable is a transformed version of the responses to question 10 of the patient interview: "Thinking now in general, what would you say the risk factors are for heart disease?" Each response was coded and a master list of all 35 responses was generated, plus one code for "Don't Know" (whgdk). For each variable, a code of "1" was given if that response was supplied, and blank otherwise. The codes are as follows:

1. family history of HD or heredity (i.e., variable whg1)
2. high blood cholesterol
3. a diet high in salt
4. high blood sugar
5. a diet low in fibre
6. being overweight
7. smoker
8. over age 45(m) or 55(F)
9. a diet high in fat
10. having high blood sugar
11. lack of activity
12. previous heart problem
13. exposure to 2nd hand smoke
14. stress
15. former smoker
16. post menopausal
17. other
18. likelihood of getting heart attack
19. risk of death
20. alcohol
21. pollution
22. exertion
23. allergies
24. stroke
25. rheumatoid arthritis
26. habits or factors that contribute to HD
27. Poor diet
28. Work
29. doesn't have any risk factors
30. too much strenuous activity
31. <code skipped in error>
32. plain bad living
33. radiation
34. something to do with the heart
35. renal problems

Appendix 3 - Programs

Creating the SAS datafile

This program was used to create a SAS file from an SPSS file provided by the client. In this example, the chart audit data was converted to a SAS file.

```
filename myfile 'C:\My Documents\Grad School\Thesis\Data files\patient data Sept 9.por';
proc convert spss=myfile out=crf.patient;
run;
```

Coding, transforming, and labeling the data

Many of the variables needed to be coded or transformed for analysis. For example, in question 10 of the patient interview, patients were asked to recall risk factors for heart disease. These were entered in up to seven variables, depending on the number of responses provided by the patient. These variables needed to be transformed into binary responses for each of 36 possible responses.

```
data;
   set crf.patient;
   if gender='m' and age>=45 then agesexr='high';else if gender='m' and age<45 then agesexr='low';
   if gender='f' and age>=55 then agesexr='high';else if gender='f' and age<55 then agesexr='low';
      array diag{37} d1-d36 ddk; array why{37} w1-w36 wdk;
      if diag21 ne ' ' then do;
         if diag21=0 then diag(37)=1;else diag(diag21)=1; end;
      if diag22 ne ' ' then do;
         if diag22=0 then diag(37)=1;else diag(diag22)=1; end;
      if why11 ne ' ' then do;
         if why11=0 then why(37)=1;else why(why11)=1;
         end;
      if why12 ne ' ' then do;
         if why12=0 then why(37)=1;else why(why12)=1;
         end;
      array risk{36} r1-r35 rdk;
      if define91 ne ' ' then do;
         if define91=0 then risk(36)=1;else risk(define91)=1;
          end;
```

```
if define92 ne ' ' then do;
   if define92=0 then risk(36)=1;else risk(define92)=1;
   end;
if define93 ne ' ' then do;
   if define93=0 then risk(36)=1;else risk(define93)=1;
    end;
if define94 ne ' ' then do;
   if define94=0 then risk(36)=1;else risk(define94)=1;
   end;
if define95 ne ' ' then do;
   if define95=0 then risk(36)=1;else risk(define95)=1;
   end;
array whatg{36} whg1-whg35 whgdk;
if whatg101 ne ' ' then do;
   if whatg101=0 then whatg(36)=1;else whatg(whatg101)=1;
    end;
if whatg102 ne ' ' then do;
   if whatg102=0 then whatg(36)=1;else whatg(whatg102)=1;
    end;
if whatg103 ne ' ' then do;
   if whatg103=0 then whatg(36)=1;else whatg(whatg103)=1;
    end;
if whatg104 ne ' ' then do;
   if whatg104=0 then whatg(36)=1;else whatg(whatg104)=1;
    end;
if whatg105 ne ' ' then do;
   if whatg105=0 then whatg(36)=1;else whatg(whatg105)=1;
   end;
if whatg106 ne ' ' then do;
   if whatg106=0 then whatg(36)=1;else whatg(whatg106)=1;
    end;
if whatg107 ne ' ' then do;
   if whatg107=0 then whatg(36)=1;else whatg(whatg107)=1;
   end;
array whatpg{36} whp1-whp35 whpdk;
if whatp121 ne ' ' then do;
   if whatp121=0 then whatpg(36)=1; else whatpg(whatp121)=1;
   end;
```

```
if whatp122 ne ' ' then do;
   if whatp122=0 then whatpg(36)=1;else whatpg(whatp122)=1;
   end;
if whatp123 ne ' ' then do;
   if whatp123=0 then whatpg(36)=1;else whatpg(whatp123)=1;
    end;
if whatp124 ne ' ' then do;
   if whatp124=0 then whatpg(36)=1;else whatpg(whatp124)=1;
    end;
if whatp125 ne ' ' then do;
   if whatp125=0 then whatpg(36)=1;else whatpg(whatp125)=1;
    end;
array active{31} act1-act30 noact;
if active1=0 then active(31)=1;else if active1 ne ' ' then active(active1)=1;
if active2=0 then active(31)=1;else if active2 ne ' ' then active(active2)=1;
if active3=0 then active(31)=1;else if active3 ne ' ' then active(active3)=1;
label whg1='Gen: Family History' whg2='Gen: High Blood Cholesterol' whg3='Gen: High Salt Diet'
   whg4='Gen: High Blood Sugar'
   whg5='Gen: Low Fibre Diet'
   whg6='Gen: Being Overweight' whg7='Gen: Smoking' whg8='Age/Gender Category'
   whg9='Gen: High Fat Diet'
   whg10='Gen: High Blood Pressure' whg11='Gen: Lack of activity'
   whg12='Gen: Previous Heart Problem' whg13='Gen: Exposure to 2nd Hand Smoke' whg14='Gen: Stress'
   whg15='Gen: Former Smoker'
   whg16='Gen: Post Menopausal'
   whg17='Gen: Other' whg18='Gen: Likelihood of Getting a Heart Attack' whg19='Gen: Risk of Death'
   whg20='Gen: Alcohol'
   whg21='Gen: Pollution' whg22='Gen: Exertion'
   whg23='Gen: Allergies' whg24='Gen: Stroke' whg25='Gen: Rheumatoid Arthritis'
   whg26='Gen: Habits of Factors that Contribute to Heart Disease' whg27='Gen: Poor Diet'
   whg28='Gen: Work' whg29='Gen: Doesn"t Have any risk factors'
   whg30='Gen: Too Much Strenuous Activity'
   whg32='Gen: Plain Bad Living' whg33='Gen: Radiation'
   whg34='Gen: Something to do with the heart' whg35='Gen: Renal Problems' whgdk='Gen: Don''t Know';
label whp1='Gen: Family History' whp2='Gen: High Blood Cholesterol' whp3='Gen: High Salt Diet'
   whp4='Pers: High Blood Sugar'
   whp5='Pers: Low Fibre Diet'
   whp6='Pers: Being Overweight' whp7='Pers: Smoking' whp8='Pers: Age/Gender Category'
```

```
   whp9='Pers: High Fat Diet'
   whp10='Pers: High Blood Pressure' whp11='Pers: Lack of activity'
   whp12='Pers: Previous Heart Problem' whp13='Pers: Exposure to 2nd Hand Smoke' whp14='Pers: Stress'
   whp15='Pers: Former Smoker' whp16='Pers: Post Menopausal'
   whp17='Pers: Other' whp18='Pers: Likelihood of Getting a Heart Attack' whp19='Pers: Risk of Death'
   whp20='Pers: Alcohol' whp21='Pers: Pollution' whp22='Pers: Exertion'
   whg23='Pers: Allergies' whp24='Pers: Stroke' whp25='Pers: Rheumatoid Arthritis'
   whp26='Pers: Habits of Factors that Contribute to Heart Disease' whp27='Pers: Poor Diet'
   whg28='Pers: Work' whp29='Pers: Doesn"t Have any risk factors'
   whp30='Pers: Too Much Strenuous Activity'
   whp32='Pers: Plain Bad Living' whp33='Pers: Radiation'
   whg34='Pers: Something to do with the heart' whg35='Pers: Renal Problems' whgdk='Pers: Don''t Know';
label r1='Family History' r2='High Blood Cholesterol' r3='High Salt Diet' r4='High Blood Sugar'
   r5='Low Fibre Diet'
   r6='Being Overweight' r7='Smoking' r8='Age/Gender Category' r9='High Fat Diet'
   r10='High Blood Pressure'
   r11='Lack of activity'
   r12='Previous Heart Problem' r13='Exposure to 2nd Hand Smoke' r14='Stress' r15='Former Smoker'
   r16='Post Menopausal'
   r17='Other' r18='Likelihood of Getting a Heart Attack' r19='Risk of Death' r20='Alcolhol'
   r21='Pollution' r22='Exertion'
   r23='Allergies' r24='Stroke' r25='Rheumatoid Arthritis'
   r26='Habits of Factors that Contribute to Heart Disease' r27='Poor Diet'
   r28='Work' r29='Doesn"t Have any risk factors' r30='Too Much Strenuous Activity'
   r32='Plain Bad Living'
   r33='Radiation'
   r34='Something to do with the heart' r35='Renal Problems' r27='Poor Diet';
label interv='Interviewer'  age='Age' gender='Gender' lang='Preferred Language'
   famdoc5='Family Doctor'
   family1='First Family Member'
   famdiag1='Diagnosis of Family Member 1' symage1='Age of onset of symptoms(fam mem1)'
   family2='Second Family Member' famdiag2='Diagnosis of Family Member 2'
   symage2='Age of onset of symptoms(fam mem2)'
   family3='Third Family Member' famdiag2='Diagnosis of Family Member 3'
   symage3='Age of onset of symptoms(fam mem3)'
   smoksta='Smoking Status' smokelen='Years Smoked' Stype1='Type Smoked' stype2='Type Smoked'
   stype1_='Amount Smoked' stype2_='Amount Smoked' mosquit='Months Quit';
label ethnic6='Ethnic Background' educat7='Education' where8='Destination on Leaving HI';
```

```
label d1='Heart Attack' d2='Angina' d3='Unstable Angina' d4= 'Post MI angina' d5=' Arrythmia'
   d6='Valve Dysfunction' d7='Valve Replacement' d8=' Artery Blockage' d9='CABG'
   d10='Cardiac Arrest'
   d11='Shortness of Breath' d12='CHF' d13='Pulmonary Edema' d14='Cardiomyopathy'
   d15='Cardiomyopathy with AICD'
   d16='Hypertension' d17='Other' d18='Angiogram' d19='PCI' d20='Unsuccessful PCI'
   d21='Heart Transplant'
   d22='Syncope' d23='Pacemaker Insertion/Check' d24='Ablation' d25='Noncardiac Chest Pain'
   d26='Noncompliance'
   d27='Stroke' d28='Loop recorder implant' d29='thoracic aneurysm repair'
   d30='Major Post-Op Complications'
   d31='Infection/Cellulitis' d32='Endocarditis' d33='Heart Problem' d34='Diabetes'
   d35='Increased Cholesterol'
   d36='Congenital Heart Defects' ddk='Don"t know';
label w1='Why? Heart Attack' w2='Why?Angina' w3='Why?Unstable Angina'
   w4='Why?Post MI angina'
   w5='Why?Arrythmia'
   w6='Why?Valve Dysfunction' w7='Why?Valve Replacement' w8='Why?Artery Blockage' w9='Why?CABG'
   w10='Why?Cardiac Arrest'
   w11='Why?Shortness of Breath' w12='Why?CHF' w13='Why?Pulmonary Edema' w14='Why?Cardiomyopathy'
   w15='Why?Cardiomyopathy with AICD'
   w16='Why?Hypertension' w17='Why?Other' w18='Why?Angiogram' w19='Why?PCI' w20='Why?Unsuccessful PCI'
   w21='Why?Heart Transplant'
   w22='Why?Syncope' w23='Why?Pacemaker Insertion/Check' w24='Why?Ablation'
   w25='Why?Noncardiac Chest Pain'
   w26='Why?Noncompliance'
   w27='Why?Stroke' w28='Why?Loop recorder implant' w29='Why?Thoracic Aneurysm Repair'
   w30='Why?Major Post-Op Complications'
   w31='Why?Infection/Cellulitis' w32='Why?Endocarditis' w33='Why?Heart Problem' w34='Why?Diabetes'
   w35='Why?Increased Cholesterol'
   w36='Why?Congenital Heart Defects' ddk='Why?Don"t know';
label info111='Info: FamDoc' sat111='Sat: Info Famdoc' info112='Info: HeartDoc'
   sat112='Sat: Info HeartDoc'
   info113='Info: HI Nurse'
   sat113='Sat: Info HI Nurse' info114='Info: Other Nurse' sat114='Sat: Info Other Nurse'
   info115='Info: FrFam'
   sat115='Sat: Info FrFam'
   info116='Info: Internet' sat116='Sat: Info Internet' info117='Info: Books' sat117='Sat: Info Books'
```

```
   info118='Info: Media'
   sat118='Sat: Info Media' info119='Info: Other' source8='Info Source (Media)'
   source9='Info Source (Other)'
   info131='PersInfo: FamDoc' sat131='Sat: PersInfo Famdoc' info132='PersInfo: HeartDoc'
   sat132='Sat: PersInfo HeartDoc' info133='PersInfo: HI Nurse'
   sat133='Sat: PersInfo HI Nurse' info134='PersInfo: Other Nurse' sat134='Sat: PersInfo Other Nurse'
   info135='PersInfo: Other' source1='Info Source (Media)'
   source1='PersInfo Source (Other)';
Label q14_1ge='Age>35/45 a RF?' q14_1p='Pers: Age >35/45' q14_2ge='History of Broken Leg a RF?'
   q14_2p='Pers: Hist Broken Leg'
   q14_3ge='Fam Hist HD a RF?' q14_3p='Pers: Fam Hist HD' q14_4ge='High Chol a RF?'
   q14_4p='Pers: High Chol'
   q14_5ge='Hist Gen Anaesth a RF?' q14_5p='Pers: Hist Gen Anaesth' q14_6ge='High Salt Diet a RF?'
   q14_6p='Pers: High Salt Diet'
   q14_7ge='High Blood Sugar a RF?' q14_7p='Pers: High Blood Sugar' q14_8ge='Low Fibre Diet a RF?'
   q14_8p='Pers: Low Fibre Diet'
   q14_9ge='Prev Surgery a RF?' q14_9p='Pers: Prev Surgery' q14_10ge='Overweight a RF?'
   q14_10p='Pers: Overweight'
   q14_11ge='Gout a RF?' q14_11p='Pers: Gout' q14_12ge='Smoking a RF?' q14_12p='Pers: Smoking'
   q14_13ge='Age >45/55 a RF?' q14_13p='Pers: Age >45/55' q14_14ge='High Fat Diet a RF?'
   q14_14p='Pers: High Fat Diet'
   q14_15ge='High BP a RF?' q14_15p='Pers: High BP' q14_16ge='Hayfever a RF?' q14_16p='Pers: Hayfever'
   q14_17ge='Lack of Phys Activity a RF?' q14_17p='Pers: Lack of Phys Activity'
   q14_18ge='Prev Heart Problems a RF?'
   q14_18p='Pers: Prev Heart Problems'
   q14_19ge='Hist of Cancer a RF?' q14_19p='Pers: Hist of Cancer'
   q14_20ge='Exp to 2nd Hand Smoke a RF?'
   q14_20p='Pers: Exp to 2nd Hand Smoke'
   q14_21ge='Hist of Chicken Pox a RF?' q14_21p='Pers: Hist of Chicken Pox' q14_22ge='Stress a RF?'
   q14_22p='Pers: Stress'
   q14_23ge='Hist Rheum Fever a RF?' q14_23p='Pers: Hist Rheum Fever' q14_24ge='Prev Smoking a RF?'
   q14_24p='Pers: Prev Smoking'
   q14_25ge='Post Menopausal a RF?' q14_25p='Pers: Post Menopausal';
label prevprob='Previous Heart Problems?' diagwht1='What diagnosis?' diagwht2='What Diagnosis'
   diagwht3='What diagnosis?'
   onset1='Age at onset?' onset2='Age at onset?' onset3='Age at onset'
   alcohol='Do you consume alcohol?';
```

Univariate Statistics

These were produced using the FREQ and MEANS procedures, examples of which follow.

```
proc freq;
  tables
     gender interv lang famdoc5 ethnic6 educat7 where8  d1-d36 ddk w1-w36 wdk r1-r35 rdk
     whg1-whg35 whgdk info111-info119 sat111-sat119 source8 source9 info131-info135 sat131-sat135 source1
     q14_1ge q14_1p q14_2ge q14_2p q14_3ge q14_3p q14_4ge q14_4p q14_5ge q14_5p q14_6ge q14_6p
     q14_7ge q14_7p q14_8ge q14_8p q14_9ge q14_9p q14_10ge q14_10p q14_11ge q14_11p q14_12ge q14_12p q14_13ge
      q14_13p q14_14ge q14_14p
     q14_15ge q14_15p q14_16ge q14_16p q14_17ge q14_17p q14_18ge q14_18p q14_19ge q14_19p q14_20ge q14_20p
     q14_21ge q14_21p q14_22ge q14_22p
     q14_23ge q14_23p q14_24ge q14_24p q14_25ge q14_25p  family1 famdiag1 family2 famdiag2 family3 famdiag3
     smoksta stype1 stype1_ stype2 stype2_ prevprob agesexr active1-active3 afreq1-afreq3 alength1-alength3
     idealex1-idealex3 ilength1-ilength3 ifreq1-ifreq3 prevprob alcohol
     whp1-whp35 whpdk act1-act30 noact;

proc means n mean std min max median lclm uclm;
        var age length3 estleng4 symage1 symage2 symage3 smokelen mosquit;
        run;
```

Transforming risk factor knowledge and predictors into binary variables

Risk factor knowledge was coded as "Y(es)", "N(o)", and "D(on't)K(now)". These variables needed to be recoded for use in logistic regression. We considered a "Y" response to reflect knowledge (=1) and "N"or "DK" to reflect lack of knowledge (=0).

```
data crf.binary;
  set crf.pat2;
  if gender='m' then male=1; if gender='m' then female=0;
  if gender='f' then female=1; if gender='f' then male=0;
  if agegrp='304' then young=1; if agegrp='304' then midage=0; if agegrp='304' then elder=0;
  if agegrp='506' then young=0; if agegrp='506' then midage=1; if agegrp='506' then elder=0;
  if agegrp='708' then young=0; if agegrp='708' then  midage=0; if agegrp='708' then elder=1;
  if edu2='1-NoH' then do;nohs=1; hsup=0;end;
  if edu2='2-HS+' then do; nohs=0;hsup=1;end;
```

```
if lang='e' then do;english=1; french=0;end;
if lang='f' then do;english=0; french=1;end;
if q14_4ge='y' then do; hicholy=1; hicholn=0;end;  if q14_4ge='n' then do; hicholy=0;hicholn=1;end;
if q14_4ge='dk' then do; hicholy=0;hicholn=1;end;
if q14_15ge='y' then do; hibpy=1; hibpn=0;end;  if q14_15ge='n' then do; hibpy=0;hibpn=1;end;
if  q14_15ge='dk' then do; hibpy=0;hibpn=1;end;
if q14_12ge='y' then do; smoky=1; smokn=0;end;  if q14_12ge='n'  then do; smoky=0; smokn=1;end;
if q14_12ge='dk' then do; smoky=0; smokn=1;end;
if q14_7ge='y' then do; hisugy=1; hisugn=0;end;  if q14_7ge='n'  then do; hisugy=0; hisugn=1;end;
if q14_7ge='dk' then do; hisugy=0; hisugn=1;end;
if q14_10ge='y' then do; ovwty=1; ovwtn=0;end;  if q14_10ge='n'  then do; ovwty=0 ; ovwtn=1;end;
if q14_10ge='dk' then do; ovwty=0; ovwtn=1;end;
if q14_17ge='y' then do; sedenty=1; sedentn=0;end;  if q14_17ge='n'  then do; sedenty=0; sedentn=1;end;
if q14_17ge='dk' then do; sedenty=0; sedentn=1;end;
if q14_6ge='y' then do; hisalty=1; hisaltn=0;end;  if q14_6ge='n'  then do; hisalty=0; hisaltn=1;end;
if q14_6ge='dk' then do; hisalty=0; hisaltn=1;end;
if q14_8ge='y' then do; lowfiy=1; lowfin=0;end;  if q14_8ge='n' then do; lowfiy=0; lowfin=1;end;
if q14_8ge='dk' then do; lowfiy=0;lowfin=1;end;
if q14_14ge='y' then do; hifaty=1; hifatn=0;end;  if q14_14ge='n'  then do; hifaty=0; hifatn=1;end;
if q14_14ge='dk' then do; hifaty=0; hifatn=1;end;
if q14_22ge='y' then do; stressy=1; stressn=0;end;  if q14_22ge='n' then do; stressy=0; stressn=1;end;
if q14_22ge='dk' then do; stressy=0; stressn=1;end;
run;
```

Calculating the risk factor index

The risk factor index (RFI) was calculated as the number of CRFs correctlyidentified. The following program created the variable RFI and produced some univariate statistics.

```
data;
   set crf.pat2;
   rfi=(q14_3g='y')+(q14_4g='y')+(q14_6g='y')+
        (q14_7g='y')+(q14_8g='y')+(q14_10g='y')+(q14_12g='y')+
        (q14_13g='y')+(q14_14g='y')+(q14_15g='y')+(q14_17g='y')+
        (q14_18g='y')+(q14_20g='y')+(q14_22g='y')+(q14_24g='y')+
        (q14_25g='y');
```

```
   label rfi='Risk Factor Index';
proc means n min mean max lclm uclm;var rfi;run;
```

Logistic Regression

Many logistic regression models were fit. We used the LOGISTIC procedure as follows, varying the modeling approach (ascending/descending/stepwise) and statistics as required.

```
data;
  set crf.binary;
/*if uniqueid^='0696190-';*/       This line removes the influential variable (with removal of /* and */).
proc logistic descending;
  class edu2 agegrp gender edu2 smoksta /param=ref;
  model lowfiy= smoksta  / rsquare details;
  proc logistic descending;
  class edu2 agegrp gender edu2 smoksta /param=ref;
  model lowfiy= gender|smoksta  / rsquare details;
  proc logistic descending;
  class edu2 agegrp gender edu2 smoksta /param=ref;
  model lowfiy= edu2 age gender  / rsquare details;
run;
```

Multiple Correspondence Analysis

For correspondence analysis, each level of each variable must be identified so that the output from the CORRESP procedure can be interpreted. In the following program, we transform the variables and perform multiple correspondence analysis, along with graphical output of the principal dimensions. This program produces the analysis depicted in Figure 11.

```
data;
  set crf.pat2;
  if q14_4ge='y' then chol='chol_y';else chol='chol_n';
  if q14_15ge='y' then hyperten='hyper_y';else hyperten='hyper_n';
  if q14_12ge='y' then smoking='smok_y';else smoking='smok_n';
  if q14_7ge='y' then bldsug='bldsug_y';else bldsug='bldsug_n';
```

```
  if q14_10ge='y' then obesity='besity_y';else obesity='obesity_n';
  if q14_17ge='y' then inact='inact_y';else inact='inact_n';
  if q14_6ge='y' then salt='salt_y';else salt='salt_n';
  if q14_8ge='y' then lowfi='lowfi_y';else lowfi='lowfi_n';
  if q14_14ge='y' then fat='fat_y';else fat='fat_n';
  if q14_22ge='y' then stress='stress_y';else stress='stress_n';
proc corresp all mca print=both outc=output;
  tables edu3 inact;
proc print data=output;
  print;
proc plot;
  plot dim1*dim2 /href=0 vref=0 box;
```

Appendix 4 – Journal Article

The following article is reproduced with the permission of the copyright owner, Lippincott Williams and Wilkins. The full citation is as follows:

Momtahan K, Berkman J, Sellick J, Kearns SA, Lauzon N. Patients' understanding of cardiac risk factors – a point prevalence study. Journal of Cardiovascular Nursing. 2004;19(1):13-20.

Journal of Cardiovascular Nursing
Vol. 19, No.1, pp 13–20 | © 2004 Lippincott Williams & Wilkins, Inc.

Patients' Understanding of Cardiac Risk Factors

A Point-Prevalence Study

Kathryn Momtahan, RN, PhD
Janet Berkman, BSc, MSc
Judith Sellick, RN, BHScN
Sharon A. Kearns, RN, BScN, CPHQ
Nancy Lauzon, RN, BA, BScN

A 1-day point-prevalence study was conducted in our 141-bed tertiary cardiac care hospital in order to determine our patients' and their significant others' level of understanding of cardiac risk factors in general and of the patients' personal cardiac risk factors. There were 3 parts to the study: patient interviews, significant other (SO) interviews, and an audit of the participating patients' charts. Of the 87 patients who were able to participate, 71 completed the interviews as did 53 significant others. From recall, only 14 patients and 11 significant others were able to define what a cardiac risk factor was ("Habits or factors that contribute to heart disease") and they were unable to identify many general risk factors. However, when given a recognition task where cardiac risk factors were interspersed with sham factors, the overall mean general knowledge score was 13.6 for patients and 13.9 for significant others out of 16. The correlation between the patients' understanding of their cardiac risk factors and the significant others' understanding of them was reasonably good ($r = 0.58$, $P < .0001$), as was the correlation between the SOs' understanding and the charts ($r = 0.58$, $P < .0001$). There was less agreement between the patients' understanding and the chart documentation of cardiac risk factors ($r = 0.36$, $P < .01$). The findings of this study have implications for patient teaching as well as for documentation of cardiac risk factors.

KEY WORDS: cardiac risk factor, point prevalence

Coronary artery disease (CAD) remains the leading cause of morbidity and mortality in Canada,[1] the United States,[2] and Europe.[3] The Canada Heart Health surveys (1985–1990) revealed that 41% of men and 33% of women aged 18 to 74 years had 2 or more of the major cardiac risk factors (smoking, high blood pressure, elevated cholesterol, physical inactivity, or obesity).[4] These surveys

Kathryn Momtahan, RN, PhD
Clinical Research Scientist, Nursing, University of Ottawa Heart Institute, Ottawa, Ontario, Canada.
Adjunct Professor, School of Nursing, University of Ottawa, Ottawa, Ontario, Canada.
Adjunct Professor, Department of Psychology, Carleton University, Ottawa, Ontario, Canada.
Janet Berkman, BSc, MSc
Graduate Student, Department of Mathematics and Statistics, University of Ottawa, Ottawa, Ontario, Canada.
Judith Sellick, RN, BHScN
Clinical Nurse Educator, Cardiac Surgery, University of Ottawa Heart Institute, Ottawa, Ontario, Canada.
Sharon A. Kearns, RN, BScN, CPHQ
Manager, Quality and Performance Measurement, University of Ottawa Heart Institute, Ottawa, Ontario, Canada.
Nancy Lauzon, RN, BA, BScN
Health Promotion Nurse, Heart Check, University of Ottawa Heart Institute, Ottawa, Ontario, Canada.

The authors would like to acknowledge the significant contribution of the nurses who participated in the study as interviewers and the Clinical Practice Committee for their dedication to clinical excellence.

Corresponding author
Kathryn Momtahan, RN, PhD, University of Ottawa Heart Institute, 40 Ruskin St, Ottawa, Ontario, Canada K1Y 4W7 (e-mail: kmomtahan@ottawaheart.ca).

focused on modifiable risk factors—risk determinants that may be modified by intervention believed to reduce the probability of specific outcomes,[5] such as mortality and morbidity. Some studies have shown that changes in lifestyle to modify these risk factors can produce a reduction in patient mortality, a reduced rate of disease progression, and occasionally even regression of atherosclerotic lesions.[6,7] Mortality from ischemic heart disease in North America has been in decline since the 1960s. Hunink et al[8] estimate that 25% of the decline in this mortality in the United States between 1980 and 1990 was due to primary prevention, 29% due to secondary prevention, and 43% due to improvements in treatment. Other nonmodifiable cardiac risk factors include age, gender, and family history of heart disease.[1,2,3]

There is some evidence that the awareness of cardiac risk factors in the general population has improved over the last 2 decades. Zerwic et al[9] cite the difference in risk factor knowledge between the study by Shekelle and Liu[10] reported in 1978 where more than half of the 617 subjects could not name any of the risk factors for heart attacks and the study by Folsom et al[11] reported in 1988 where 70% of the adults correctly identified at least 1 of the 3 main modifiable risk factors (smoking, high blood pressure, and high cholesterol or fat in the diet). Zerwic et al asked 105 patients, hospitalized for myocardial infarction or for coronary angiography with the diagnosis of coronary artery disease, open-ended questions regarding cardiac risk factors. They found that 79% of the patients named at least 1 of 3 modifiable risk factors (smoking, hypertension, elevated cholesterol) but only 7% identified all 3.

The Clinical Practice Group at the University of Ottawa Heart Institute, a 141-bed regional cardiac care hospital, identified the understanding of cardiac risk factors in the hospital patient population as a research priority. As a result, a 1-day point-prevalence study was conducted during a 12-hour period midweek. All willing and able in-hospital patients and their significant others were interviewed. Audits of the participating patients' charts were done in the 2-month period following the day of interviews.

Method

Sample

Seventy-one patients of the 121 patients who were hospitalized at our facility between 0900 and 2100 hours on a Wednesday in June 2002 were interviewed. Twenty hospital beds were closed on this day. Twenty-six patients were excluded from the study. The exclusion criteria included patients who were not literate in English or French or did not have someone to translate for them, patients who were confused and unable to participate because of their condition (for instance, were mechanically ventilated or deemed too ill by the interviewer to participate), and those who were assessed to be too close in time to a procedure or operation to participate. Eight patients who may have been willing to participate were missed because of timing issues (for example, they were discharged suddenly or were otherwise unavailable). After exclusions and missed patients, 87 patients were eligible to participate in the study. Thirteen patients declined to participate and 3 patients decided not to continue after they had signed the consents. This resulted in 71 patients participating in the study, an 82% response rate; 89% of the patients fell into a high-risk age/gender group, that is, females older than 55 years and males older than 45 years.

Setting

The study was conducted at the University of Ottawa Heart Institute (UOHI), a teaching hospital that provides bilingual French/English regional cardiac services to a catchment area in excess of 1.8 million residents. Patients in all units except the operating rooms, the cardiac recovery room, the cardiovascular ICU, and the cardiac catheterization laboratories who were not excluded for the previously cited reasons, were asked to participate in the study along with their significant other (SO). A *significant other* was defined as the person that the patient thought would be most involved with their care and recovery once they left the hospital. The SO was usually a spouse but could be another family member or a close friend. The decision was made to survey the SOs as well as the patients for 2 reasons: family and friends may be instrumental in helping patients recover and modify their lifestyles once they return home,[5] and we wanted to determine the congruence in the answers given by the patients and their SOs regarding the patients' cardiac risk factors.

Procedure

There were 3 parts to the study: the patient interviews, the SO interviews, and the chart audits. All of the interviewers and chart auditors were nurses employed at the UOHI.

Interviews

Twelve nurses interviewed 71 patients and the majority of the 53 SOs during the 12-hour period. Some of the SOs were not available in hospital to be interviewed. They were later contacted by telephone. The patient interviews and the SO interviews were

conducted separately to avoid bias. The interviewers participated in two 3-hour training sessions prior to the day of interview. Interrater reliability testing was done to achieve 85% agreement on the information gathered from mock interviews.

Patients and their SOs were asked to define what a cardiac risk factor is. They were also asked to identify cardiac risk factors in general and to identify what the patient's personal cardiac risk factors were. Following these open-ended questions, patients and their SOs were presented with what was described as a list of possible cardiac risk factors. Sixteen were in fact risk factors, whereas 9 of these were sham risk factors. They were asked to identify whether each item on the list was a cardiac risk factor in general; if the answer was yes, they were asked if this was a risk factor that they personally had.

Chart Audits

The completed charts of the 71 patients who were interviewed were audited by 10 auditors over the 2-month period following the point-prevalence interview day once the chart was received in medical records with a discharge diagnosis assigned. Eighty-two percent of the charts were audited by 4 nurses who had also participated in the interviewing of patients and their SOs. To ensure that the charts were reviewed consistently, the first day of chart auditing was done in a group by these 4 auditors. Because the data gathered for the chart audits was less ambiguous than the information gathered by the interviews, no formal interrater reliability assessment was done.

Results

Patient Interviews

Sixty-two percent of the patients interviewed were male and 38% were female. Ninety-six percent of the patients were Caucasian. Sixteen percent of the patients had completed university, 18% had completed college, 24% had completed high school, and 42% had not completed high school. Ninety-four percent of the patient interviews were conducted in English and 6% in French. The median length of hospital stay at the time of the interview was 4 days (range 0–150 days). Patients' estimates of how long they would be in hospital ranged from 1 to 171 days, with the median being 6 days. Eighty-three percent of patients were planning to go home from hospital. Another 10% were planning to go to the home of a friend or family member. The remainder were returning to regional hospitals or were going to a convalescent home.

Although the patients were asked both "Why are you in hospital?" and "What is your diagnosis?", the responses they gave were almost identical. Table 1 lists the summary of the diagnostic categories. Although there were many different answers, most of the categories could be collapsed. For instance, the category *Acute coronary syndrome* encompasses angina, unstable angina, heart attack, post-MI angina, cardiac arrest, and artery blockage.

TABLE 1 Summary of Patients' Responses to the Question "What is Your Diagnosis?"

Diagnostic Category	Number	(%)	Male (%)	Female (%)	Mean Age (y)
Arrhythmia	6	8.5	9.1	7.4	71
Heart failure	3	4.2	6.8	0	79
Surgical procedure	13	18.3	22.7	11.1	63
Noninvasive procedure	1	1.4	2.3	0	58
Acute coronary syndrome	33	46.5	45.4	48.2	64
Postoperative complications	3	4.2	2.3	7.4	55
Other	7	9.9	6.8	14.8	63
Don't know	5	7	4.6	11.1	50
Total	71	100	100	100	

Patients were asked to define "cardiac risk factor". There were 14 correct responses ("habits or factors that contribute to heart disease") and 28 patients gave no response. There were 31 mentions of specific risk factors, as well as statements along the line of "likelihood of getting a heart attack"(10) or "risk of death"(2). There were also vague responses: "poor diet"(8), and "plain bad living"(3). Table 2 presents the answers patients gave when asked about cardiac risk factors either unprompted or prompted.

What is noteworthy about Table 2 is the huge differences between the unprompted and prompted responses. Had we only asked unprompted questions regarding the patients' understanding of cardiac risk factors in general as well as their own personal risk factors, we would have been under the impression that patients had very little understanding of cardiac risk factors.

One of the most striking pieces of information regarding cardiac risk factors in this population is that 17% of patients stated that they were smokers, with 61% reporting being smokers in the past, indicating that 78% of these inpatients were either smokers or former smokers and only 22% had never smoked. These figures differ substantially from those found on the charts: 22% smokers, 38% past smokers, and 34% who had never smoked. Six percent of the charts did not contain any information regarding the patient's smoking history.

TABLE 2 Patients' Responses to Questions Regarding Cardiac Risk Factor

	General					Personal				
		Prompted					Prompted			
Risk Factors	Unprompted	Yes	No	Don't Know	Total	Unprompted	Yes	No	Total	
Modifable cardiac risk factors										
High blood cholesterol	12	68	0	3	71	10	38	23	61	
High salt diet	3	61	6	4	71	1	18	39	57	
High blood sugar	10	59	5	7	71	9	19	35	54	
Low fiber diet	1	35	20	16	71	0	8	22	30	
Overweight	15	68	2	1	71	5	29	34	63	
Smoking	50	68	2	1	71	20	21	45	66	
Former smoker	0	57	9	5	71	6	14	37	51	
High fat diet	16	66	2	2	70	4	26	38	64	
High blood pressure	5	69	0	2	71	6	31	31	62	
Lack of physical activity	22	61	7	3	71	8	26	33	59	
Stress	20	65	4	2	71	11	33	29	62	
Nonmodifiable cardiac risk factors										
Family history of heart disease	22	68	2	1	71	12	48	17	65	
Male >45 y, female >55 y	0	57	11	3	71	0	48	6	54	
Previous heart problem	0	66	0	5	71	3	31	33	64	
Exposed to secondhand smoke	0	63	5	3	71	1	27	28	55	
Sham risk factors										
Male >35 y, female >55 y		36	27	8	71		30	3	33	
History of broken leg		5	61	5	71		1	4	5	
Having had a general anaesthetic		7	41	23	71		1	5	6	
Having had previous surgery		16	40	15	71		6	7	13	
Gout		17	26	28	71		3	14	17	
Hay fever		6	35	30	71		2	4	6	
History of cancer		14	34	23	71		1	9	10	
Having had chicken pox		4	38	29	71		1	2	3	
History of rheumatic fever		38	15	17	70		1	35	36	

TABLE 3 Patients' Information Sources on Cardiac Risk Factors in General

Information Source	Received (%)	Satisfied (%)
Media	66	84
Family doctor	47	81
Heart Institute doctor	40	75
Friends and family	37	92
Heart Institute nurse	36	92
Books	36	91
Internet	19	77
Other nurse	11	67
Other	29	95
Pamphlet	15	
Dietician	6	
Other doctor	5	
Lecture	2	
MI class	2	

The media was the most frequently mentioned source of information on cardiac risk factors in general, with two thirds of patients obtaining information about general risk factors from those sources (Table 3). Fewer patients had received information on personal risk factors from various sources, including health care professionals (Table 4).

When asked about the type of physical exercise they did, 53% of the patients said that they walked, although only 6% (4 patients) indicated that they walked vigorously. Eighteen percent listed exercise that could be considered aerobic in nature as the form of exercise they performed and 21% said that they did not exercise. Fifty-one percent of patients stated that they consumed alcohol, compared with 44% of the SOs that said the patient consumed alcohol. Alcohol consumption was reported in 29% of the charts.

Significant Other Interviews

Twelve interviewers interviewed 53 SOs. Eight patients said that they did not have an SO. One SO was unable to participate because he/she spoke neither French nor English. The remainder were refusals, 1 of which was by a patient who did not want the SO involved. This resulted in an overall response rate of 85%. One subject was unable to complete the interview, and so some questions in the latter part of the interview reflect this lower response rate ($n = 52$). Ninety-two percent of the SO interviews were conducted in English and 8% in French. Of 53 SOs, 68% were female, 26% were male, and 6% were unknown (coded as "friend"). Spouses comprised 72% of SOs, and a child (son or daughter) was SO for 17%. Other relations were mother, grandmother, and sister. SOs were asked to define cardiac risk factor. There were 11 correct responses ("habits or factors that contribute to heart disease") and 16 SOs with no response or a "don't know" response. There were 31 mentions of specific risk factors, as well as statements along the line of "likelihood of getting a heart attack" (8) or "something to do with the heart" (4). There were also vague responses: "poor diet" (6), "work" (1), "exertion/strenuous activity" (2), and "plain bad living" (3).

TABLE 4 Patients' Sources of Information on Their Personal Cardiac Risk Factors

Information Source	Received (%)	Satisfied (%)
Family doctor	35	78
Heart institute doctor	34	73
Heart institute nurse	29	95
Other nurse	12	100
Other	26	94
Friends and family	10	
Pamphlet	7	
Newspaper	3	
Television	3	
Diet clinic	2	
Dietician	2	

Chart Audits

The mean age of the 71 patients who were interviewed and whose charts were audited was 63 years, ranging from 31 to 86 years. The median length of stay was 8 days, with the range being 1.5 to 184. Mean body mass index (BMI) was 27.9 (range 18.1–38.0). Forty-nine percent of the patients were overweight (BMIs between 25 and 30), and 24% were obese (BMIs > 30). Fifty-six percent of patients had hypertension (either measured on admission, stated in their history, or mentioned elsewhere on the chart). Thirty-nine percent of charts documented a family history of heart disease. Fourteen percent stated that there was no such history, and 45% of charts provided no information either way.

Risk Factor Knowledge

By summing the number of correct answers to the 16 risk factor questions, we created an index that can proxy the patients' general knowledge of cardiac risk factors. A second index measures self-reported personal risk factors.

General Risk Factor Knowledge (Patient)

The overall mean general knowledge score was 13.6 (out of a possible 16.) Males scored on average 13.2 and females scored 14.2. The test for equivalence of means yielded $P = .08$. While this difference is not statistically significant at the usual $\alpha = .05$ level, this

TABLE 5 Percentages of Personal Cardiac Risk Factors Identified by the Patients, Their Significant Others, and in the Charts

Personal Risk Factors	Patients (n = 71)	SO (N = 52)	Chart (N = 71)	(%) Agreement: Patient + SO (n = 52)	Patient + Chart (n = 71)	SO + Chart (n = 52)
Family history of HD	68	69	41	75	56	48
High blood cholesterol	54	56	51	69	62	67
High blood sugar	27	21	35	85	70	83
Being overweight	41	38	20	62	68	65
Being a smoker	30	23	17	83	79	81
High blood pressure	44	52	62	71	65	75
Lack of physical exercise	37	48	23	69	61	69
Previous heart problems	44	44	64	73	63	65
Stress	46	58	7	60	49	46
Former smoker	20	25	8	73	83	73
Age	68	71	89	63	76	81

Note: The last 3 columns present the percent agreement between the following pairs: the patients and their SOs, the patient and the chart, and the SO and the chart.

is a suggestive result particularly in light of the gender differences in SO scores (below). Those with high school or higher education scored significantly higher than those who did not complete high school (14.2 vs 12.8; P = .012.) Examining college-educated and university-educated individuals separately from those with high school education did not improve knowledge score. There was no difference in mean score between groups of patients with and without previous heart problems (13.5 vs 13.7). Mean knowledge scores by age group were as follows: 30 to 49 years: 12.9; 50 to 69 years: 14.0; 70 to 89 years: 13.2.

General Risk Factor Knowledge (Significant Other)

When SOs were asked about general cardiac risk factors (CRFs), their overall mean score was 13.9. Excluding the 3 "friend" SOs whose gender is not known, female SOs identified more risk factors than did the males (P = .001). Of the 52 patient/SO pairs we had data for, 25 of the SOs scored higher than the patient, 9 scores were identical, and 18 of the patients scored higher than the SO. Using recognition rather than a recall task, most of the items that were cardiac risk factors were correctly identified by patients as well as their SOs.

Personal Risk Factors

Of 14 possible risk factors that were looked for on the chart, between 0 and 10 were identified for individual patients, with a mean of 4.5. The mean number of personal risk factors identified by patients was 6.0 (min = 0, max = 13). Significant others identified 6.4 risk factors on average (min = 0, max = 13). The percentages of reported personal risk factors by patients, their SOs, and the charts, are listed in Table 5.

If we just consider the number of risk factors identified (the score), we can compute the correlation between the various sources of personal and general risk factors. These correlations are presented in Table 6. There is significant correlation between the number of risk factors identified in the chart, by the patient (personal) and by the SO (personal).

It is clear that, at least on average, patients and their SOs identified more risk factors than what was found on the charts. These correlations, recalculated using only the modifiable cardiac risk factors, are presented in Table 7.

TABLE 6 Correlations Between Cardiac Risk Factors Identified by Patients, Their SOs, and the Patients' Health Care Team (From Chart audits)

		Patient		SO	
	Chart	General	Personal	General	Personal
Chart	1	0.12	0.37†	−0.06	0.58§
Patient (general)		1	0.35†	0.32*	0.26
Patient (personal)			1	0.04	0.58§
SO (general)				1	0.17
SO (personal)					1

*P < .05.
†P < .01.
‡P < .001.
§P < .0001.

TABLE 7 Correlations Between a Core Set of 7 Modifiable Cardiac Risk Factors Identified by Patients, Their SOs, and the Patients' Health Care Team (From chart audits)

	SO		
	General	Personal	Chart
Patient (general)	0.21		
Patient (personal)		0.59*	0.51*
SO (personal)			0.63*

Note: Modifiable risk factors (7) include: high blood cholesterol, high blood sugar, being overweight, being a smoker, high blood pressure, lack of physical exercise, and Stress.
*$P < .0001$

Discussion

On the recommendation of our hospital's Clinical Practice Committee, comprised of nurses from all units of the hospital, we set out to discover the level of understanding of cardiac risk factors among our patient population and their significant others. A 1-day point prevalence study was conducted. There are several findings that are noteworthy.

Most patients were unable to define cardiac risk factor or to identify many cardiac risk factors spontaneously. However, given a multiple-choice format, they were able to correctly identify almost all of them and correctly identified the items that were inserted as sham factors and were not cardiac risk factors. Only 7% of the patients of Zerwic et al[9] were able to identify 3 main modifiable risk factors (smoking, high blood pressure, and high cholesterol). If we compare this to our patient population, only 4.2% of the patients and 3.8% of the SOs could identify all 3 when asked unprompted. However, prompted, 93.0% of the patients and 96.2% of the SOs identified all 3.

The study found good agreement between knowledge of general and personal risk factors between patients and their SOs. This is an encouraging result and suggests that patients have significant others who are aware of what the patient's cardiac risk factors are. This may be helpful in supporting the patients when they are trying to modify their risk factors, although this cannot be assumed.[12]

The chart audit revealed a lack of consistent data related to cardiac risk factors and planned interventions. This finding highlighted the need for a multidisciplinary review of documentation associated with risk factors, an important first step in being able to measure institutional compliance with evidence-based standards of care and improved outcomes for patients. Fonarow et al[13] implemented a cardiac hospitalization atherosclerosis management program (CHAMP) aimed at initiating secondary protection measures based on the American Heart Association guidelines for secondary prevention while patients were in hospital.[14] Two years later, this group was able to demonstrate significant and positive effects of the initiative on patient medication compliance, LDL cholesterol levels, and improved clinical outcomes when compared with patients 2 years prior to the implementation of the CHAMP program.

The BMI calculations indicated that 73% of the inpatient population on this day is either overweight or obese. These numbers are higher than those found in the general population[1] but might be expected in an older population of patients with cardiovascular disease. This points to a need for ongoing educational strategies for diet and weight reduction.

The data in this study indicate that patients and their significant others have a good general knowledge of cardiac risk factors when prompted. However, the fact that they needed prompting in order to recall this information argues in favor of a more structured and collaborative approach to teaching patients and their families about cardiac risk factors. This research identified the patient group most in need of specific information regarding their personal risk factors as being males aged between 30 and 40 years. This is consistent with previous literature. It has been shown that the patient population that undergoes percutaneous coronary intervention (PCI) is less likely to make lifestyle modifications, indicating to some authors that these patients may feel they are cured of their disease after this relatively unobtrusive intervention.[15,16] The danger of this perception is that without lifestyle modification, the risk of further cardiovascular progression remains significant in a group that might reap the most benefits from lifestyle modifications.[17]

One of the major challenges to providing effective in-hospital education programs is the decreasing number of days this patient population is actually in hospital. Myocardial infarction (MI) patients and patients who undergo PCIs are in hospital for just 1 to 4 days. In-hospital education is provided during times of stress, and because of this the amount of information these patients are able to absorb is variable.[18,19] Patient education and access to resources is key to any form of patient participation framework.[20–23] In addition to compliance with prescribed medication, it is important that patients with coronary artery disease learn to care for themselves and modify their modifiable cardiac risk factors. This can be facilitated by the development of explicit standards for risk factor assessment and documentation. In addition, appropriate patient education materials, in-hospital as well as home resources (eg, hospital-based rehabilitation programs, Web sites, community

programs) are required to support patients in risk factor modification.

REFERENCES

1. Heart and Stroke Foundation and Health Canada. *The Changing Face of Heart Disease and Stroke in Canada 2000*. Ottawa, Canada; 1999.
2. American Heart Association. *Heart Disease and Stroke Statistics—2003 Update*. Dallas: American Heart Association; 2002.
3. EUROSPIRE I and II Group. Clinical reality of coronary prevention guidelines: a comparison of EUROSPIRE I and II in nine countries. *Lancet*. 2001;357:995–1001.
4. Joffres MR, Ghadirian P, Fodor JG, Petrasovits A, Chockalingam A, Hamet P. Awareness, treatment, and control of hypertension in Canada. *Am J Hypertens*. 1997;10(10, pt 1):1097–1102.
5. Jairath N. *Coronary Heart Disease & Risk Factor Management*. Philadelphia: Saunders; 1999.
6. Haskell WL, Alderman EL, Fair JM, et al. Effects of intensive multiple risk factor reduction on coronary atherosclerosis and clinical cardiac events in men and women with coronary artery disease: the Stanford Coronary Risk Intervention Project (SCRIP). *Circulation*. 1994;89(3): 975–990.
7. National Institutes of Health (NIH), National Heart, Lung and Blood Institute. *Educational Strategies to Prevent Prehospital Delay in Patients at High Risk for Acute Myocardial Infarction*. Bethesda, Md: NIH; 1997. NIH Publication 97-3787.
8. Hunink MG, Goldman L, Tosteson AN, et al. The recent decline in mortality from coronary heart disease, 1980–1990: the effect of secular trends in risk factors and treatment. *JAMA*. 1997;277:535–542.
9. Zerwic JJ, King K, Wlasowicz GS. Perceptions of patients with cardiovascular disease about the causes of coronary artery disease. *Heart Lung*. 1997;26(2):92–98.
10. Shekelle RB, Liu SC. Public beliefs about causes and prevention of heart attacks. *JAMA*. 1978;240:756–758.
11. Folsom AR, Sprafka M, Luepker RV, Jacobs DR. Beliefs among black and white adults about causes and prevention of cardiovascular disease: the Minnesota heart survey. *Am J Prev Med*. 1988;4:121–127.
12. Daugherty J, Saarmann L, Riegel B, Sornborger K, Moser D. Can we talk? Developing a social support nursing intervention for couples. *Clin Nurse Spec*. 2002;16(4):211–218.
13. Fonarow GC, Gawlinski A, Moughrabi S, Tillisch JH. Improved treatment of coronary heart disease by implementation of a cardiac hospitalization arteriosclerosis management program (CHAMP). *Am J Cardiol*. 2001;87:819–822.
14. Ryan TJ, Antman EM, Brooks NH, et al. ACC/AHA guidelines for management of patients with acute myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Acute Myocardial Infarction). *J Am Coll Cardiol*. 1999;34: 890–911.
15. Gaw-Ens B, Laing G. Risk factor reduction behaviours in coronary angioplasty and myocardial infarction patients. *Can J Cardiovasc Nurs*. 1994;5(1):4–12.
16. Cronin S, Freeman L, Gingy R, Drake D. Recovery after percutaneous transluminal coronary angioplasty: assessment after discharge. *Crit Care Nurse*. 2000;20(2):70–76.
17. Menotti A, Lanti M. Coronary risk factors predicting early and late coronary deaths. *Heart*. 2003;89:19–24.
18. Wingate S. Post-MI patients' perceptions of their learning needs. *Dimens Crit Care Nurs*. 1990;9(1):33–44.
19. Wang WW. The educational needs of myocardial infarction patients. *Prog Cardiovasc Nurs*. 1994;9(4):28–36.
20. Brezynskie H, Pendon E, Lindsay P, Adam A. Identification of the perceived learning needs of balloon angioplasty patients. *Can J Cardiovasc Nurs*. 1998;9(2):8–14.
21. Czar M, Engler M. Perceived learning needs of patients with coronary artery disease using a questionnaire assessment tool. *Heart Lung*. 1997;26(2):109–117.
22. Turton J. Importance of information following myocardial infarction: a study of the self-perceived information needs of patients and their spouse/partner compared with the perceptions of nursing staff. *J Adv Nurs*. 1998;27(4): 770–778.
23. Cahill J. Patient participation—a review of the literature. *J Clin Nurs*. 1998;7(2):119–128.

Appendix 5 – Correspondence Analysis Notation and Process

This notation is excerpted from Friendly(2000).

Let $\mathbf{N}=\{n_{ij}\}$ be the observed data, i=1,…,I and j=1,…,J.

Let $\mathbf{P}=\{\mathbf{N}/n\}$, the matrix of joint cell probabilities, or correspondence matrix.

In the example under consideration (**Table 13**), n=71,

$$N=\begin{bmatrix} 7 & 1 & 4 & 0 \\ 17 & 12 & 7 & 7 \\ 6 & 4 & 2 & 4 \end{bmatrix} \text{ and } P=\begin{bmatrix} 0.10 & 0.01 & 0.06 & 0.00 \\ 0.24 & 0.17 & 0.10 & 0.10 \\ 0.08 & 0.06 & 0.03 & 0.06 \end{bmatrix}.$$

Let $\underline{r}=\sum_j p_{ij}$, the row margin or row mass of **P** and similarly, let $\underline{c}$ be the column margin (column mass) of P.

$$\underline{r}=\begin{bmatrix} 12 \\ 43 \\ 16 \end{bmatrix} \text{ and } \underline{c}=\begin{bmatrix} 30 \\ 17 \\ 13 \\ 11 \end{bmatrix}$$

Let $\mathbf{D}_r$ and $\mathbf{D}_c$ be the diagonal matrices with $\underline{\mathbf{r}}$ and $\underline{\mathbf{c}}$ on their diagonal.

$$D_r=\begin{bmatrix} 12 & & \\ & 43 & \\ & & 16 \end{bmatrix} \text{ and } D_c=\begin{bmatrix} 30 & & & \\ & 17 & & \\ & & 13 & \\ & & & 11 \end{bmatrix}$$

Then $R=D_r^{-1}P=\left\{\frac{n_{ij}}{n_{+j}}\right\}$ is the matrix of row conditional probabilities or row profiles.

Similarly, **C** is the matrix of column conditional probabilities or column profiles. In the example

$$R = \begin{bmatrix} 0.58 & 0.08 & 0.33 & 0.00 \\ 0.40 & 0.28 & 0.16 & 0.16 \\ 0.37 & 0.25 & 0.13 & 0.25 \end{bmatrix} \text{ and } C = \begin{bmatrix} 0.23 & 0.06 & 0.31 & 0.00 \\ 0.57 & 0.71 & 0.54 & 0.64 \\ 0.20 & 0.23 & 0.15 & 0.36 \end{bmatrix}$$

The singular value decomposition of P is given by

$$P = AD_{\lambda}B^{T}$$

where

$$D = \begin{bmatrix} \lambda_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \lambda_M \end{bmatrix} \text{ and } \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$$

and $\boldsymbol{A}$ is the IxM matrix where $\boldsymbol{AD_r^{-1}A^T=I}$ and $\boldsymbol{B}$ is the JxM matrix where $\boldsymbol{BD_C^{-1}B^T=I,}$ and $\lambda_1,\ldots,\lambda_M$ are the M eigenvalues (or singular values). M is equal to the total number of dimensions, the minimum of {(I-1),(J-1)}. In this example, M=2 (that is, min(2,3)).

Define $F = D_r^{-1}AD_{\lambda}$ and $G = D_c^{-1}BD_{\lambda}$ such that $F^T D_r F = D_{\lambda}$ and $G^T D_c G = D_{\lambda}$. The matrices F and G are the row and column profiles expressed with respect to their own principal axes. In this example,

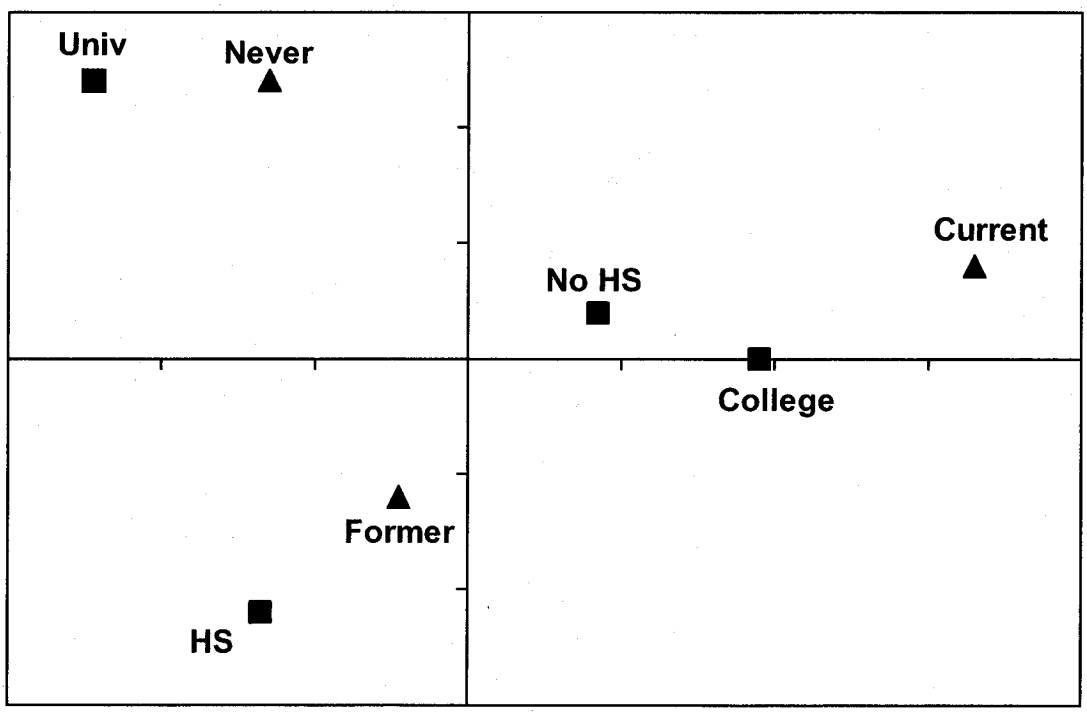
$$D_{\lambda} = \begin{bmatrix} 0.31 & \\ & 0.07 \end{bmatrix},$$

$$F = \begin{bmatrix} 0.66 & 0.04 \\ -0.09 & -0.06 \\ -0.26 & 0.12 \end{bmatrix}$$

$$\text{and } G = \begin{bmatrix} 0.17 & 0.02 \\ -0.27 & -0.11 \\ 0.38 & 0.00 \\ -0.49 & 0.12 \end{bmatrix}.$$

By plotting the rows of F and G (Figure A1), we can visualize the relationship between the row and column vectors in 2-dimensional space. In this example, the dimensionality of the problem is 2, therefore we see 100% of the variability.

Figure A1
CA Plot of Smoking Status and Education



In this example, we note that the points for university graduates and never smokers are close together. Similarly, high school grads and former smokers lie in the same quadrant. Finally, those with no high school education, or a college education, and current smokers lie in the same quadrant.

[1] Momtahan K, Berkman J, Sellick J, Kearns S, Lauzon N. *atients' understanding of cardiac risk factors: a point-prevalence study.* J Cardiovasc Nurs 2004;19(1):13-20.

[2] International Classification of Diseases (ICD), 9th Revision, Codes 390-459.

[3] Statistics Canada. Statistical Report on the Health of Canadians. Sept. 16, 1999.

[4] Langille DB, Joffres MR, MacPherson KM, Andreou P, Kirkland SA, MacLean DR. *Prevalence of risk factors for cardiovascular disease in Canadians 55 to 74 years of age: results from the Canadian Heart Health Surveys, 1986-1992.* CMAJ. 1999;161(8 Suppl):S3-9.

[5] Leventhal, H., Meyer, D. Nerenz, D. The common sense representation of illness danger. In Rachman, S. (ed.), Contribution to Medical Psychology. Pergamon Press, Oxford, vol. 2, pp. 7-30. 1980

[6] Wilson PW, D'Agostino RB. *Prediction of coronary heart disease using risk factor categories.* Circulation. 1998;97:1837-1847.

[7] Shekelle RB, Liu SC. *Public beliefs about causes and prevention of heart attacks* JAMA. 1978;240:756-758

[8] Folsom AR, Sprafka M, Luepker RV, Jacobs DR. *Beliefs among black and white adults about causes and prevention of cardiovascular disease: the Minnesota heart survey.* Am J Prev Med. 1988;4:121-127

[9] Zerwic JJ, King K, Wlasowicz GS. *Perceptions of patients with cardiovascular disease about the causes of coronary artery disease.* Heart Lung. 1997;26(2):92-98.

[10] Kirkland SA, MacLean DR, Langille DB, Joffres MK, MacPherson KM. *Knowledge and awareness of risk factors for cardiovascular disease among Canadians 55 to 74 years of age: results from the Canadian Heart Health Surveys 1986-1992.* CMAJ 1999;161(8 Suppl): S10-S16.

[11] Health Promotion and Disease Prevention: Provisional Data from the National Health Interview Survey: United States, January–June 1985. Hyattsville, Maryland: National Center for Health Statistics, 1986; DHHS Pub. No (PHS) 86-1250.

[12] Wong, ND; Black HR, Gardin JM. *Preventive Cardiology.* New York NY: McGraw-Hill, 2000:388.

[13] Hosmer DW, Lemeshow S. *Applied Logistic Regression (2nd Ed.).* New York NY: John Wiley and Sons, 2000.

[14] Friendly, M. *Visualizing Categorical Data.* Cary NC: SAS Publications, 2000.

[15] Clausen, Sten-Erik. *Applied Correspondence Analysis.* Thousand Oaks CA: Sage Publications, 1998.

[16] Fienberg SE. *The Analysis of Cross-Classified Categorical Data.* Cambridge MA: The MIT Press., 1977.

[17] Allison PD. *Logistic Regression Using the SAS® System: Theory and Application.* Cary NC: SAS Institute Inc., 1999.

[18] Little RJA, Rubin DB. *Statistical Analysis with Missing Data (2nd Ed.).* Hoboken NJ: John Wiley and Sons, 2002.

[19] Fleiss JL. *Statistical Methods for Rates and Proportions* (2nd Ed). New York NY: John Wiley and Sons, 1981.

[20] Friedman LM, Furberg CD, DeMets DL. *Fundamentals of Clinical Trials (3rd Ed).* New York: Springer, 1998.

[21] Dempster AP et al. *Maximum likelihood from incomplete data via the EM algorithm (with discussion).* JRSS. 1977;39:1-38.