Environmental Factors Effects on Asthma in the National Population Health Survey

Environmental Factors Effects on Asthma in the National Population Health Survey

By

Rongfang Gu, B. Eng.

A Project

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

March, 2003

MASTER OF SCIENCE (2003)
(Statistics)

McMaster University
Hamilton Ontario

TITLE: **Environmental Factors Effects on Asthma in the National Population Health Survey**

AUTHOR: **Rongfang Gu**

SUPERVISOR: **Dr. Román Viveros-Aguilera**

NUMBER OF PAGES: **x, 104**

Abstract

Asthma is one of the most common respiratory diseases in Canada. It not only brings pain to more than 7.8% of Canadians but it also costs millions of dollars every year.

In this project we study the influence that environmental factors have on asthma based on the data from the National Population Health Survey (NPHS) conducted by Statistics Canada. A descriptive analysis is done first to get an initial understanding of the environmental factors' effects on asthma. Then we do a χ^2-test to test for the homogeneity of asthma distribution across the levels of each environmental factor. Most of the factors included in our study are significant except those representing whether living in the metropolitan areas of Montreal or Vancouver, the number of persons living in the household, and whether there are small children in the household.

For ordinal variables we test for trend on asthma prevalence. The trend tests indicate that there are significant trends between asthma and most of the ordinal factors except for a few, including the number of bedrooms in the household and the number of cigarettes smoked daily by the daily smokers.

Then odds ratio and relative risk analyses are done to obtain statistical insights on the relative risk of the factors. The result shows that living in Nova Scotia, Ontario and Quebec, urban areas, engaged in finance, community services, personal service, young, attend physical activities, born in Canada, white, single or widowed, separated, divorced, start smoking early, do not own the dwelling living in, female adult, male children, and overweight and underweight adult, are all contributing factors for asthma.

To model and investigate the joint effect of factors on asthma prevalence, we use logistic and log-linear regression models. To avoid collinearity problems, a reduced number of predictors is used. The results suggest that environmental factors have a significant joint influence on asthma prevalence.

Acknowledgements

I am very grateful to have this opportunity to express my deep appreciation to my supervisor, Dr. Román Viveros-Aguilera, for his inspiring direction and support throughout the entire process of this project. Dr. Román Viveros-Aguilera is a patient and understanding supervisor. I am very fortunate to have his great guidance.

Many thanks to Mr. Vivek Jadon from the Data/Text Center at Mills Library, McMaster University for his enthusiasm and help during the initial stages of my project.

I thank Dr. Peter Macdonald and Dr. Lehana Thabane for their careful reading of my project and for their insightful suggestions that led to many improvements in the final version of my project.

I am also grateful to my professors, staff and my fellow graduate students in the Department of Mathematics and Statistics of McMaster University. It was them who made my studies enjoyable and worthwhile.

I am extremely thankful to my parents for their encouragement and support. A special thanks to my husband for his deep love.

TABLE OF CONTENTS

Chapter 3: Odds Ratio and Relative Risk Analysis

Chapter 4 Loglinear and Logistic Regression Analysis

List of Tables

List of Figures

Chapter 1: The National Population Health Survey (NPHS): Data and Study Objectives

1.1 Introduction

The National Population Health Survey (NPHS) is designed to collect information related to the health of the Canadian population. The survey began in 1994, and conducted and managed by Statistics Canada by-yearly. It is composed of three component parts: the survey of households, the survey of institutions and the survey of the North (Yukon and the Northwest Territories with the principal exclusion of populations on Indian Reserves, Canadian Forces Bases and some of the most northerly remote areas of the Territories.). The household component includes household residents in all provinces, with the exclusion of populations on Indian Reserves, Canadian Forces Bases and some remote areas in Québec and Ontario. In each household, some limited information was collected from all household members (general component) and one person in each household was randomly selected for a more in-depth interview (health component). The 1996-97 NPHS was collected mainly by telephone. For that reason, an effort has been made to make the questions easier to read (for the interviewers) and easier to understand (for the respondents) by using more colloquial wording where possible. Without revamping sections, screens have been redesigned, response categories have been shortened and frequent responses in "other specify" have become itemized categories.

The source data for our study come from the 1996-97 NPHS Public Use Microdata Documentation. It should be noted that the "public use" microdata files differ in a number of important respects from the survey "master" files held by Statistics Canada. These differences are the result of actions taken to protect the anonymity of individual survey respondents through suppression of individual values, variable grouping, and variable capping.

The principle behind estimation in a probability sample such as the NPHS is that each person in the sample has a sampling weight. WT66 is the principal weight variable on the Health File and should be used for analysis of most variables. For many analysis techniques (for example linear regression, logistic regression, analysis of variance), a method exists that can make the application of standard packages more meaningful. If the weights on the records are rescaled so that the average weight is 1, then the results produced by the standard packages will be more reasonable since they take into account the unequal probabilities of selection. The rescaling can be accomplished by using in the analysis a weight equal to the original weight divided by the average of the original weights for the sampled units (people) contributing to the estimator in question. This approach was taken in this project.

As indicated earlier, the data came from Statistics Canada, one of the most respectable institutions in the country for conducting surveys. Since I did not have access to the original surveys, it was not possible to verify the reliability of the data.

1.2 Case and Variable Selection, Sub-Data Description

The NPHS data are stored into two different data sets. The information collected about all household members is stored on the General File. Generally, in each household, one person was selected to answer a more in-depth questionnaire related to health. These data are stored on the Health File.

Each record on the Health File corresponds to the household of the selected respondent. This file carries all data collected pertaining to the selected respondent, e.g. all the variables from the General File belonging to the selected respondent and all variables resulting from the in-depth questionnaire, such as smoking, stress, mental health, etc. Since the Health File contains more information, we chose it for our main study.

The data analyzed in this project are the 1996-1997 NHPS data obtained in CDs from the Data/Text Center, Mills Library, McMaster University. The main file is the Health File which is stored as a text file named H356.txt. Using the SAS program listed in the Appendix B we unzipped and restored the data set. The Health File data contains 81,804 records and 944 variables. The 944 variables relate to the following characteristics: alcohol dependence, attitudes towards parents, blood pressure, chronic conditions, coping, drug use, demographic and household variables, dental visits, education, emergency services, eye examination, flu shots, geographic identifiers, general health, health care utilization, health information, health status, HIV, height and weight, injury, income, insurance, labor force, mental health, physical activity, physical check-up, province, restriction of activity, repetitive strain, road safety, socio-demographics, sexual health, smoking, social support, health services, training and UV exposure, two-week disability, violence and personal safety, women's health, and sample weights.

Because of its known high prevalence, particularly in children, asthma has been an intensively studied disease. Rees and Kanabar (2000) report that ethnicity, family history, rural / urban environment, body weight, indoor environment (particularly for children), vigorous exercise, occupation, emotional factors, pollution, weather and air quality all have an influence on asthma. The studies reported vary in their focus of analysis and methods, and usually based on data from specific parts of the world. Manning (1993) discuses the effects that different forms of exercise have on asthma. One of his conclusions is that more than 50% of asthmatic subjects experience progressively less and less airway narrowing with repeated exercise challenges on the same day. Exercise induced asthma is also studied by Tan and Spector (2002). Other studies focus on the effect on asthma of more restricted environments such as school (Leickly, 2003) and urban/rural areas (Maffei *et al.*, 2001). Since our main objective is to study the environmental effects on asthma prevalence, we need not to include all the 944 variables from the survey. After reviewing the survey questionnaire, the meaning of each variable and our study objectives, 35 variables are selected for study. So a sub data set from the main source data (Health File) is used in our study to do the analysis. The sub data still has 81,804 records but with only the 35 variables which are of interest in our study. The variables included in our sub data can be divided into groups that describe outdoor environment, indoor environment, working environment, physical exercise and air exchange amount, social demographics and whether in a smoking environment. We also include variables that describe basic characteristics of respondents in our sub data.

The selected variables are listed in Table 1.1. For details about variable meaning, coding, range and universe please see Appendix A.

Table 1.1 List of variable names and meaning descriptions.

	Variable Code	Description
Main dependent variable	CCC6_1C	Has asthma (Do you have asthma diagnosed by a health professional? Yes or No)
Outdoor environment variables	PRC6_CUR	Province of residence
	GE36GCMA	Derived 1991 Census Metropolitan Area - grouped
	GE36GHLR	Derived health areas – grouped (26 groups)
	GE36GHRO	Derived health areas – grouped (33 groups)
	GE36GURB	Derived rural and urban area – grouped
Working environment variables	LFC6GO21	Occupation Codes for main job – grouped (21 groups)
	LFC6GI13	Industry Codes for main job – grouped (13 groups)
Physical exercise, air exchange amount (respiratory system)	PAC6DEE	Derived energy expenditure (1 decimal point)
	PAC6DLEI	Derived participant in leisure physical activity
	PAC6DFM	Derived monthly freq. of physical activity lasting >15min.
	PAC6DFR	Derived frequency of all physical activity
	PAC6DFD	Derived participation in daily phys. Activities > 15 min.
	PAC6DPAI	Derived physical activity index
Demographic variables	SDC6GCB	Country of birth – grouped
	SDC6GRAC	Derived race or color – grouped
	DHC6GMAR	Marital status – grouped

	Variable Code	Description
Whether in smoking environment	SMC6_3	Age started smoking daily – daily smoker
	SMC6_4	Number of cigarettes smoked each day – daily smoker
Indoor environment variables	DHC6_OWN	Dwelling owned by household member
	DHC6GBED	Derived number of bedrooms in dwelling – grouped
	DHC6GHSZ	Derived household size – grouped
	DHC6GLE5	Derived persons <= 5 years old in household – grouped
	DHC6G611	Derived persons 6 to 11 years old in hhld – grouped
	DHC6GECF	Derived household type – grouped
	DHC6DLVG	Derived living arrangements of the selected respondent
Basic variables of respondents' characteristics	HWC6GHT	Height - adults and children - grouped
	HWC6GSW	Derived standard weight - grouped
	HWC6G3KG	Weight in kilograms - grouped
	HWC6GBMI	Derived Body Mass Index (1 decimal place) - grouped
	DHC6_SEX	Gender
	DHC6GAGE	Age - grouped
	CCC6_1H	Has chronic bronchitis or emphysema
	DGK6_1	Takes ventolin or other inhalants
Sampling weight	WT66	Sampling weight for selected respondent

1.3 Study Objectives

The overall objectives of this project are to study the influence of environmental factors on asthma. More concretely, our main aims can be described as follows.

(a) To assess individually which of the factors studied have a significant influence on asthma prevalence. This will be accomplished by using appropriate graphical methods and by a Chi-square test.

(b) Detect trends between asthma prevalence and ordinal categorical and non categorical factors. The Cochran-Armitage test will be used.

(c) For factors that have a significant influence on asthma prevalence from (a), study the relevant odds ratios and relative risks by constructing 2×2 tables.

(d) Relate the factors in a joint manner to the asthma prevalence through logistic regression models and do the associated estimation and inference. First, the model will be applied to the whole data. Then the data will be split into infants, children and adults. Each of these groups will be randomly partitioned into three subsets, one for variable selection, the second one for model fitting and the third are for cross-validation.

1.4 Descriptive Analysis of Data

1.4.1 Response Variable CCC6_1C

Table 1.2 displays the sample frequency distribution for the response variable in this project, CCC6_1C (Has Asthma). The corresponding general population frequency distribution data are from the 1996 national census of Canada. All information in Table 1.2 can be found in the data dictionary of 1996-1997 NPHS survey. Most respondents have unambiguous answers on whether they have asthma or not. The respondents who answer "Don't know", "Refusal" and "Not stated" account for only 50 out of 81,804, that is, 0.061% of all the respondents. To simplify the analysis we will delete those respondents who do not answer "Yes" or "No" on whether they have asthma. So in the sub data we study, there are 81754 cases. Note that among the Yes / No respondents, 7.64% and 7.81% have asthma in the sample and in the population, respectively.

Table 1.2 Frequency distribution for variable CCC6_1C (Has Asthma).

Variable Name	CCC6_1C	**Length**	1	**Position**	165	
Question Name	CHR-Q1					
Concept	Has asthma					
Question	Do you have asthma diagnosed by a health professional?					
Universe	All respondents					
Note						
Content			**Code**		**Sample**	**Population**
YES			1		6,242	2,236,139
NO			2		75,512	26,394,976
DON'T KNOW			7		38	7,859
REFUSAL			8		5	523
NOT STATED			9		7	2,239
		Total			81,804	28,641,736

1.4.2 Independent Variables

In the following we review the sub data from the seven groups of variables described above. Most of the graphs were generated using Splus and Excel.

The NPHS survey shows that the overall asthma prevalence rate in Canada is 7.81%. Table 1.3 depicts the asthma rate per province while Figure 1.1 displays the corresponding histogram. However, the prevalence rate distribution among provinces and health areas is different . The highest occurs at the province of Nova Scotia, which is as high as 8.43%. The second highest is Ontario, which is 8.04%. The lowest ones are Newfoundland and New Brunswick, with 5.76% and 6.6%, respectively.

Table 1.3 Asthma rate per province.

Province	Asthma Prevalence (%)	Percentage of people with asthma in each province to the total number of asthma suffers from all provinces (%)
Nova Scotia	8.43	3.38
Ontario	8.04	38.95
Quebec	7.98	25.16
P.E.I.	7.78	0.46
British Columbia	7.76	12.79
Alberta	7.52	9.17
Manitoba	7.39	3.58
Saskatchewan	6.95	2.95
New Brunswick	6.6	2.15
Newfoundland	5.76	1.41
Whole Canada	7.81	100

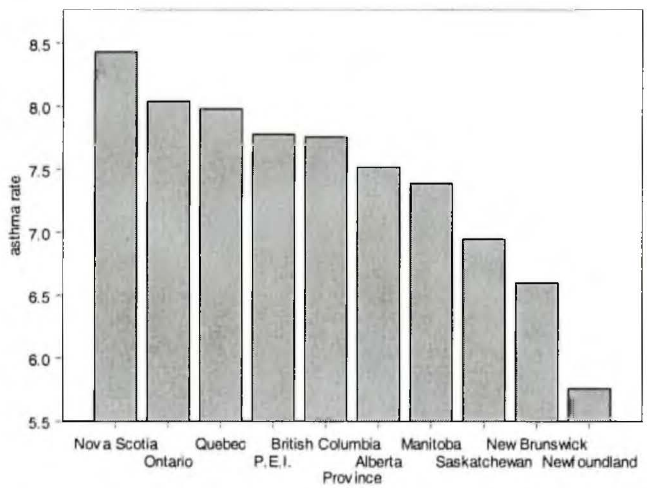


Figure 1.1 Asthma rate for the Canadian provinces.

Since the provinces vary dramatically in size, so do the total count of asthma sufferers. These variations, which appear in column 3 of Table 1.3 are plotted in Figure 1.2. Note these figures are the gross percentages, they have not been adjusted for provinces size.

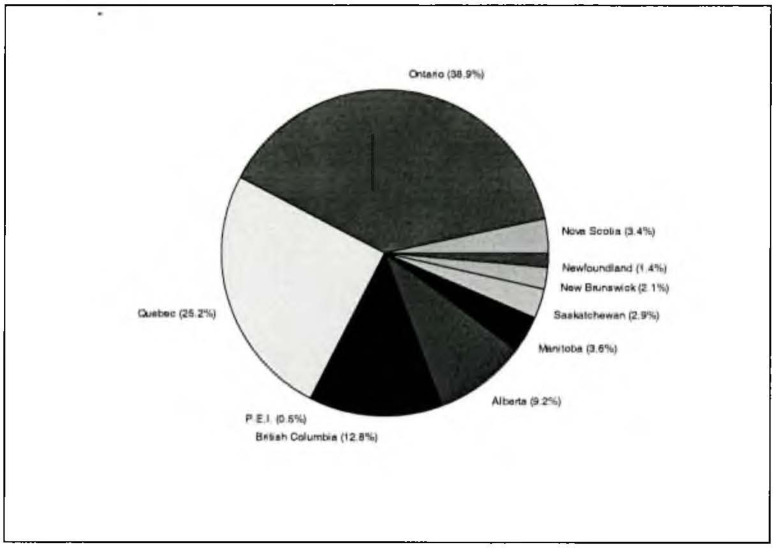


Figure 1.2 Pie chart of the percentage of asthma sufferers per province.

Variable GE36GURB (whether living in urban or rural area) provides another view of asthma distribution. People living in urban areas have an asthma rate of 7.88%, whereas people living in rural areas have the smaller rate of 6.52%. This partly means that a rural environment is a little "healthier" than an urban environment for asthma sufferers.

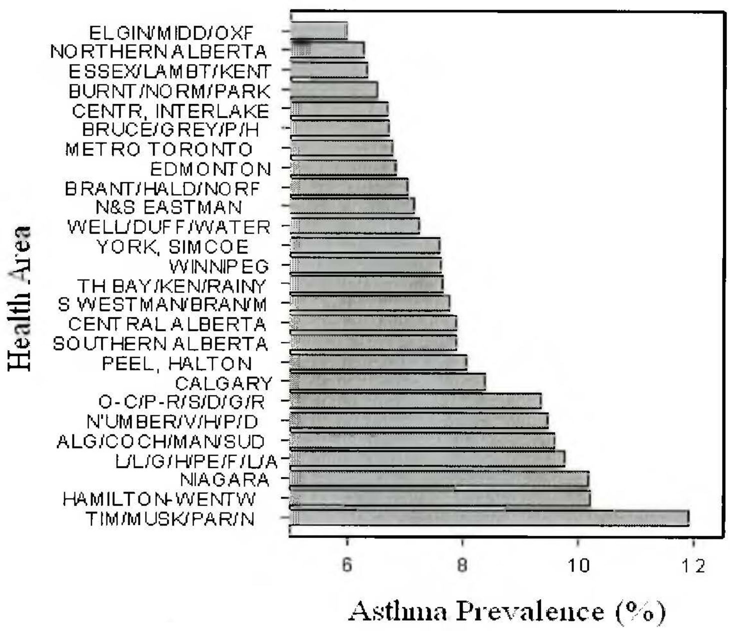


Figure 1.3 Asthma rate among health areas for Ontario, Manitoba and Alberta.

Some provinces such as Ontario, Manitoba and Alberta provide larger financial and human resources thus more detailed surveying is done. The three provinces account for more than a half of the total population. The three provinces are divided into 26 health areas; the histogram in Figure 1.3 displays the asthma rates among the 26 areas.

It is clear that the asthma rate is not uniform across areas. From the frequency table we note that the asthma rate is as low as 5.96% for the area of ELGIN, MIDDLESEX and OXFORD and as high as 11.93% for the area of TIMISKAM., MUSKOKA, PARRY SOUND and NIPISS.

Occupational asthma is a disease characterized by variable air flow limitation or airway hyper-responsiveness due to causes and conditions attributable to a particular occupational environment and not to stimuli encountered outside the workplace (Bernstein *et al.* 1993).

Variable LFC6GO21 has 21 groups (Figure 1.4) and LFC6GI13 has 13 groups both relate to occupation. From Figure 1.4 it is clear that occupation is a potentially influential variable on asthma rate.

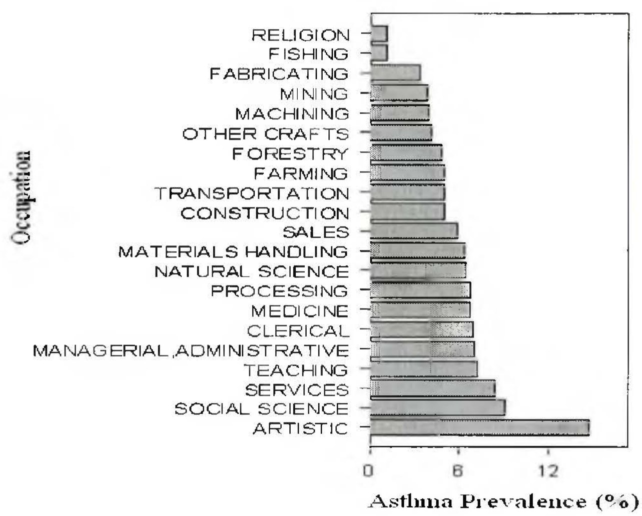


Figure 1.4 Asthma prevalence among occupations.

The ARTISTIC, SOCIAL-SCIENCE, and SERVICES occupations show the highest asthma rates; FISHING and RELIGION have the lowest rate. Note, however, that some of these occupation groups are much smaller than others.

Physical exercise or physical activities can cause or worsen asthma symptoms (Ree and Kanabar, 2000, p. 16; Tan and Spector, 2002). A few people seem to get asthma attacks only when they run or take other exercise. This 'exercise-induced asthma' is especially a problem for young people. Exercise is just one of many things which show that the air passages are irritable in asthma. Cold air, tobacco smoke (e.g. in a pub), emotional stress, infections (such as colds and flu), sulphur dioxide (used as a preservative in soft drinks and wine) can all be triggers of asthma or asthma attacks (AAIR, 2003). The survey data also support this viewpoint. People who answered yes to the question of whether they participate in leisure time activity have an overall higher asthma rate.

For PAC6DFD which describe whether respondents participate in daily physical activities more than 15 minutes, people who answer yes have an asthma rate of 8.59% while people answer no have an asthma rate of 6.71%. The same pattern exists in other variables such as physical activity index.

For the demographic environment variables such as SDC6GCB (country of birth), SDC6GRAC (race or color), DHC6GMAR (marital status), the frequency tables generated by SAS show that respondents born in Canada have a higher asthma rate (8.6%) than people born in USA or Europe (5.4%), respondents born in Asia have an asthma rate as low as 1.7%, and white people have a higher asthma rate (8.01%) than other races (6.12%).

Marital status is another demographic variable we are interested in. For the three category groups in this survey, we find that single persons have the highest asthma rate of 10.67%, followed by widow/separate/divorced person at 7.94% and married/living with a common law/partner at only 5.44%.

Smoking is one of the risk factors often thought to increase the risk of developing asthma. For individuals with asthma, smoking can make symptoms worse. In our study we find individuals who smoke daily have an asthma rate of 8.01%, higher than the overall rate of 7.81%. For those who smoke, variable SMC6_3 provides information about the starting age of their smoking habit. From Figure 1.5 we see that individuals with asthma start smoking relatively earlier than healthy ones.

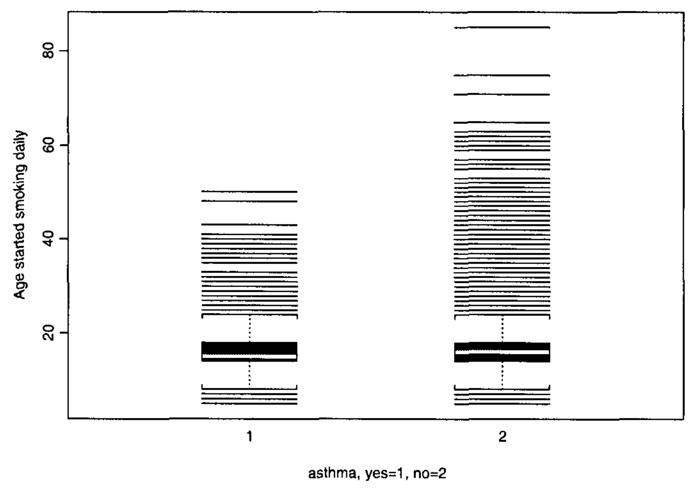


Figure 1.5 Boxplot of respondents' age of started smoking daily.

But for those who smoking daily, there seems to be little difference on the number of cigarettes smoked daily between the asthma group and non-asthma group (see Figure 1.6).

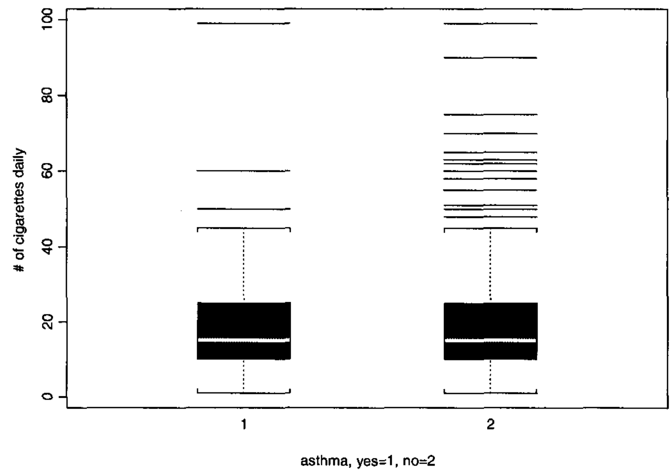
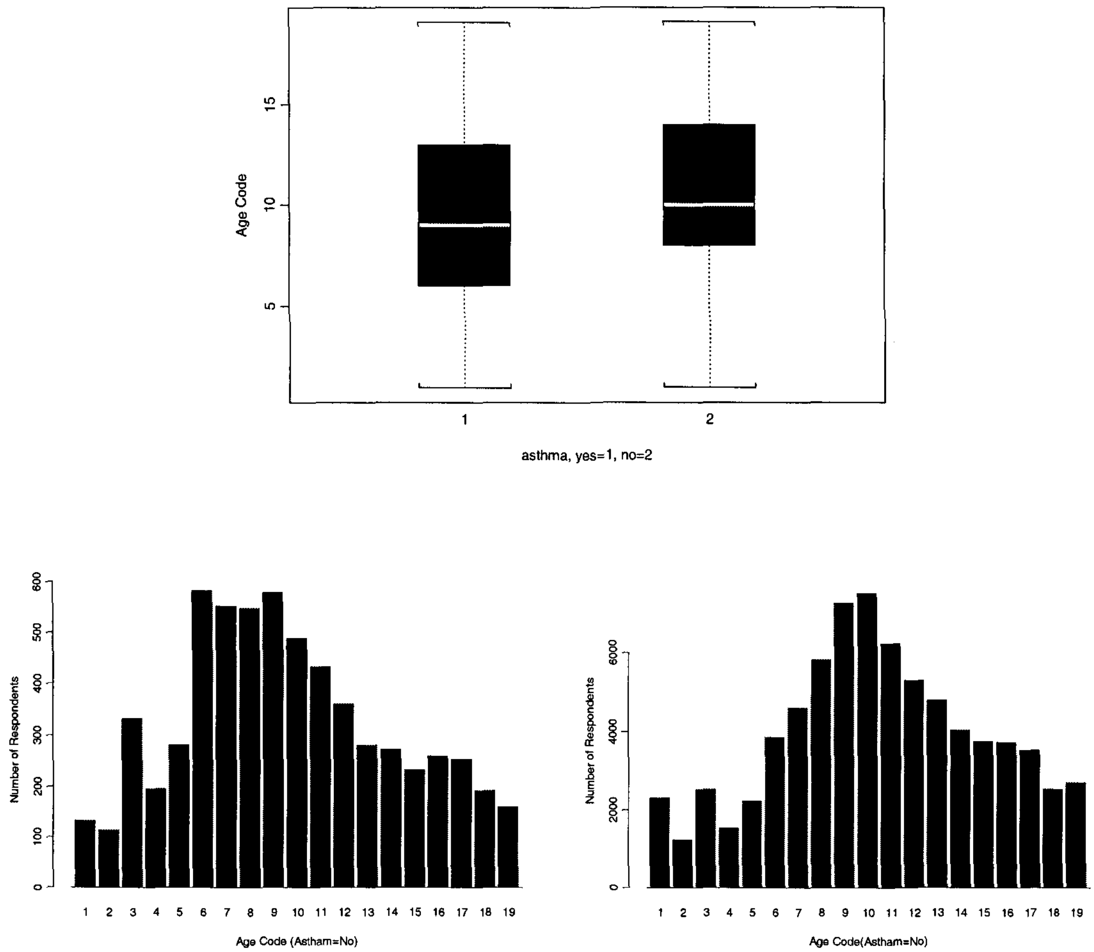


Figure 1.6 Boxplot of the number of cigarettes smoked per day.

Indoor air is another important factor that influences asthma prevalence. Research results from the Canadian Human Activities Pattern Survey (CHAPS) indicate that adults in Canada spend about 90% of their time indoors (Reports on Respiratory Disease in Canada of CHAPS, 2003, p. 27). As multiple concomitant exposures may heighten sensitivities, a combination effect is important to consider when indoor air exposures are suspected to be a cause of illness. Our data show that people who own the dwelling they live in have a lower asthma rate of 7.08%, by contrast the rest have asthma rate of 9.81%. The number of bedrooms also shows some effect on asthma rate. People living in dwellings with 1, 2, 3, 4 bedrooms have asthma rate of 8.91%, 8.41% 7.38% and 7.48%, respectively. People in households with 1, 2, 3, 4, 5 persons living in has asthma rate of 7.73%, 7.58%, 7.71%, 8.18% and 7.76%, respectively. Grouping people by whom they are living with reveals that single parents with children living together have the highest asthma rate of 12.68%.

The last issue we explore here is about the effect of basic characteristics of respondents on asthma prevalence.



Age Code	Age Range		Age Code	Age Range		Age Code	Age Range
1	0 TO 3 YEARS		8	25 TO 29 YEARS		15	60 TO 64 YEARS
2	4 TO 5 YEARS		9	30 TO 34 YEARS		16	65 TO 69 YEARS
3	6 TO 9 YEARS		10	35 TO 39 YEARS		17	70 TO 74 YEARS
4	10 TO 11 YEARS		11	40 TO 44 YEARS		18	75 TO 79 YEARS
5	12 TO 14 YEARS		12	45 TO 49 YEARS		19	80 YEARS OR OLDER
6	15 TO 19 YEARS		13	50 TO 54 YEARS			
7	20 TO 24 YEARS		14	55 TO 59 YEARS			

Figure 1.7 Age distribution of respondents who have asthma and who do not have asthma.

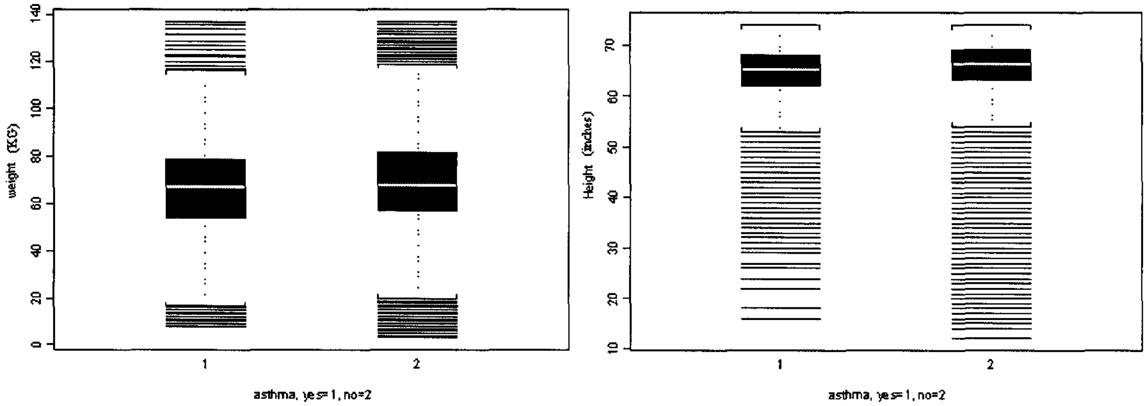


Figure 1.8 Boxplots of respondents' weight and height.

From Figure 1.7 we see that individuals with asthma have a relative young age. SAS results show that asthma prevalence is different by gender. From boxplots in Figure 1.8 we see that Weight and Height seem to have little effect on asthma prevalence. A look at the distribution by gender shows that females have a higher asthma prevalence than males, there respective prevalence rates are 8.47% and 7.16%. Our calculations also show that individuals with the airway diseases of chronic bronchitis or emphysema the asthma prevalence soar as high as 33.87%.

Chapter 2: Contingency Table Analysis for the Selected Variables in the NPHS

2.1 Contingency Table Analysis

Contingency tables are used to classify possible values of two or more variables. In each cross classified cell the number of observations is reported. They can be viewed as joint frequency distributions of variables.

We focus on the case of two variables: a row variable (R) and a column variable (C). Let R vary by a row index i, where i=1,2,…I, and let C vary by a column index j, where j=1,2,…J. It is customary to let R denote the explanatory variable and C the outcome variable. We denote the cell frequency of the ith row and the jth column by f_{ij} and the respective expected frequency by F_{ij}. See Table 2.1.

Table 2.1 Contingency table structure.

		Respondent variable				
		Level 1	Level 2	…	Level J	Total
Explanatory Variable	Level 1	f_{11} (F_{11})	f_{12} (F_{12})	…	f_{1J} (F_{1J})	f_{1+} (F_{1+})
	Level 2	f_{21} (F_{21})	f_{22} (F_{22})	…	f_{2J} (F_{2J})	f_{2+} (F_{2+})
	……	……	……	…	…	……
	Level I	f_{I1} (F_{I2})	f_{I2} (F_{I2})	…	f_{IJ} (F_{IJ})	f_{I+} (F_{I+})
	Total	f_{+1} (F_{+1})	f_{+2} (F_{+2})	…	f_{+J} (F_{+J})	f_{++} (F_{++})

A widely used test statistic for testing the independence model is the Pearson's χ^2 statistic. Denote by $\hat{F}_{ij}$, the estimated expected frequency under independence, $\hat{F}_{ij} = f_{i+} f_{+j} / f_{++}$, then Pearson's χ^2 statistic can be shown to be:

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}(\hat{F}_{ij} - f_{ij})^2 / \hat{F}_{ij} \ . \tag{2.1}$$

It can be proved (see Agresti, 1990, p. 44) that this statistic has an asymptotically chi-squared distribution with degrees of freedom equal to (I-1)(J-1) when the hypothesis of independence is true.

In the contingency table analysis we also can use the likelihood ratio method. This method leads to another widely used statistic in contingency table analysis, the log likelihood test statistic G^2 given by:

$$G^2 = -2\log Q = 2\sum_{i=1}^{I}\sum_{j=1}^{J} f_{ij} \log(f_{ij} / \hat{F}_{ij}) \ . \tag{2.2}$$

G^2 is asymptotically distributed as χ^2 under the independent hypothesis. The degrees of freedom can be calculated as the difference between the parameters under the unrestricted model (IJ-1) and the number of parameters under the restricted model (I-1)+(j-1) resulting in (I-1)(J-1) (Powers and Xie, 2000, pp. 88-105).

2.2 Application of Contingency Table Analysis to the NPHS Data

Using the programs listed in Appendix C we test for independence between the status of asthma (CCC6_1C) and each of the variables selected for study. The results are listed in Appendix B. Examining the P-values for both the Pearson method and the likelihood ratio method (LLR), all the variables show a significant association with asthma except for: GE36GCMA (Derived 1991 Census Metropolitan Area - grouped), DHC6GHSZ (Derived household size - grouped) and DHC6GLE5 (Derived persons <= 5 years old in household - grouped). This suggests that most of the variables we selected from the source data set are potentially useful to explain asthma prevalence. That is, there is a significant difference in distribution between asthma and non-asthma groups for these variables.

Note that a total of 30 Chi-square statistical tests are reported in Appendix B. This means the overall probability of making the Type I error may be large. Note, however, that the P-values reported are very small for the most part (<0.0001) due to the large sample sizes. Thus, the single-test level of significance α could be made sufficiently small (much smaller than 5%) to offset the overall Type I error probability.

To see this in detail, consider m independent tests with a common significance level α. The overall significance level is:

$P_m(\alpha)$=P(rejecting independence in at least one test | all tables are independent) .

Clearly $P_m(\alpha)=1-(1-\alpha)^m$. Table 2.2 shows that, for m=30 tests, $P_m(\alpha)<0.05$ if $\alpha<0.0017$.

Table 2.2 Values of the overall Type I error $P_m(\alpha)$

for individual significance levels α and m=30.

α	$P_m(\alpha)$
0.0001	0.0030
0.0005	0.0149
0.0017	0.0498
0.0052	0.1448

2.3 Test for Trend

Many of the variables contained in the data set are ordinal scaled. We are interested in whether there is a discernible trend in the level of association between the factor and the asthma prevalence rate. The trend test used in this section addresses this question and is very useful for two-way contingency tables with binomial response variables. The test is sensitive to the linearity between response variable and experimental variables and detects trends that would otherwise not be noticed by more crude methods. It is based upon the regression coefficient for the weighted linear regression of the binomial proportions on the scores of the levels of the explanatory variable. See Margolin (1988), Agresti (1990), and Stokes *et al.* (2000).

For I × 2 tables with ordered rows, using the linear probability model, for each row let $y_{1|I}$ denote the probability of response 1, and let $p_{1|i}$ denote the sample proportion, i = 1, . . . , I. Let { x_i } be scores assigned to the rows. For the linear probability model

$$y_{1|i} = \alpha + \beta x_i \quad . \tag{2.3}$$

the ordinary least squares fit gives the prediction equation

$$\hat{y}_{1|i} = p_{+1} + b(x_i - \bar{x}) \quad . \tag{2.4}$$

where

$$\bar{x} = \frac{\sum (n_i + x_i)}{n} \quad , \quad b = \frac{\sum [n_i + (p_{1|i} - p_{+1})(x_i - \bar{x})]}{\sum [n_i + (x_i - \bar{x})^2]} \quad . \tag{2.5}$$

The Cochran-Armitage asymptotical standard normal statistic for trend is then given by

$$z = \left\{ \frac{b^2}{p_{+1}p_{+2}} \sum n_i + (x_i - \overline{x})^2 \right\}^{1/2} \quad . \tag{2.6}$$

When the linear probability model holds, the statistic z, tests for a linear trend in the proportions. The trend test may give strong evidence regarding positive or increasing linear trends, constant or stable trends over time, and negative or decreasing trends. SAS computes one-sided and two-sided p-values for the trend test (See SAS/STAT User's Guide Version 7 and Manitoba Centre for Health Policy and Evaluation, 1999). When the test statistic is greater than its expected value of zero, SAS computes the right-sided p-value, which is the probability of a larger value of the statistic occurring under the null hypothesis. A small right-sided p-value supports the alternative hypothesis of increasing trend in column 1 probability from row 1 to row I. When the test statistic is less than or equal to zero, SAS computes the left-sided p-value. A small left-sided p-value supports the alternative of decreasing trend (see SAS/STAT User's Guide Version 7).

Table 2.2 summarizes the z value obtained from the SAS analysis (programs is listed in the Appendix):

Table 2.2 Results of trend test.

Variable	Cochran-Armitage Trend Test
DHC6GAGE (Age – grouped)	Statistic (Z) -24.3017 One-sided Pr < Z <.0001 Two-sided Pr > |Z| <.0001
DHC6GBED (Derived number of bedrooms in dwelling – grouped)	Statistic (Z) 1.2591 One-sided Pr > Z 0.1040 Two-sided Pr > |Z| 0.2080
PAC6DFR (Derived frequency of all physical activity)	Statistic (Z) -5.9025 One-sided Pr < Z <.0001 Two-sided Pr > |Z| <.0001

Variable	Cochran-Armitage Trend Test
SMC6_4 (Number of cigarettes smoked each day - daily smoker)	Statistic (Z) -3.8102 One-sided Pr < Z <.0001 Two-sided Pr > |Z| 0.0001
SMC6_4(<=15) (Number of cigarettes smoked each day<=15 cigarettes)	Statistic (Z) -1.1955 One-sided Pr < Z 0.1159 Two-sided Pr > |Z| 0.2319
SMC6_4(>15) ((Number of cigarettes smoked each day>15 cigarettes)	Statistic (Z) -2.0695 One-sided Pr < Z 0.0193 Two-sided Pr > |Z| 0.0385
HWC6GHT (Height - adults and children – grouped)	Statistic (Z) -15.3412 One-sided Pr < Z <.0001 Two-sided Pr > |Z| <.0001
HWC6G3KG (Derived standard weight – grouped-all respondents)	Statistic (Z) -15.0593 One-sided Pr < Z <.0001 Two-sided Pr > |Z| <.0001
HWC6G3KG (Derived standard weight – grouped-for respondents weights more than 70kg)	Statistic (Z) 2.3580 One-sided Pr > Z 0.0092 Two-sided Pr > |Z| 0.0184
HWC6GBMI (Derived Body Mass Index – grouped, respondents aged 20~64 years old)	Statistic (Z) 6.0396 One-sided Pr > Z <.0001 Two-sided Pr > |Z| <.0001
SMC6_3 (Age started smoking daily)	Statistic (Z) -7.2516 One-sided Pr < Z <.0001 Two-sided Pr > |Z| <.0001

From Table 2.3 we see that at the α=5% level of significance there are decreasing trends in age, height, and age at smoking start to the prevalence of getting asthma. There is no trend between number of bedrooms, energy expenditure and asthma. Test to all respondents show that there is a decreasing trend between the number of cigarettes smoking daily to the prevalence of getting asthma. But if we look at the data we find that for those who smoke less than 15 cigarettes everyday there is no trend, trend only exists in respondents who smoke more than 15 cigarettes everyday. Body mass index shows that for respondents who are 20 to 64 years old there is an increasing trend between the value of the index and asthma prevalence. Study of the variable of respondents' weight also shows this trend for the same age group. There is also increasing trend in derived standard weight.

In summary up the above results, we conclude that the younger the respondents the higher the asthma prevalence. Smoking has an adverse effect on respondents: the early they start smoking daily the larger the asthma prevalence. The decreasing trend between cigarettes smoked daily and asthma prevalence seems peculier at first. But we should consider the fact that the elder and longer a person smokes, the larger the potential the person smokes more cigarettes daily. For those who smoke more there may be some adaptation developed which helps them to defend against asthma. It is also true that people who suffer asthma tend to smoke less if they do have the smoking habit. Asthma tends to be suffered by those who are young and whose respiratory systems are not getting used to the smoking environment.

Chapter 3: Odds Ratio and Relative Risk Analysis

3.1 Odds Ratio and Relative Risk

It is helpful to describe the chances that a binary response variable leads to a success in terms of the odds of that event. The odds of success is defined to be the ratio of the probability of a success to the probability of a failure. Thus if p is the true success probability, the odds of success is p/(1-p).

Table 3.1 A hypothetical 2×2 table.

	Yes	No	Total	Proportion of Yes
Group 1	n_{11}	n_{12}	n_{1+}	$\hat{p}_1=n_{11}/n_{1+}$
Group 2	n_{21}	n_{22}	n_{2+}	$\hat{p}_2=n_{21}/n_{2+}$
Total	n_{+1}	n_{+2}	n	

For the 2×2 table displayed in Table 3.1, the odds ratio of interest is the one that compares the odds of the Yes proportion for Group 1 to the odds of the Yes proportion for Group 2. Specifically, denote the odds ratio as O_R, thus its estimate is

$$\hat{O}_R = \frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = \frac{n_{11}n_{22}}{n_{12}n_{21}}, \qquad (3.1)$$

where p_1 is the asthma probability in group 1 and p_2 is the asthma probability in group 2 whose estimates are $\hat{p}_1$ and $\hat{p}_2$.

The odds ratio ranges from 0 to infinity. When O_R is 1, there is no association between the row variable and the column variable. When O_R is greater than 1, group 1 is more likely than group 2 to have a Yes response and vise versa. It is convenient to take log of odds ratio, thus we get:

$$f = \log\{\hat{O}_R\} = \log\left\{\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)}\right\} = \log\{\hat{p}_1/(1-\hat{p}_1)\} - \log\{\hat{p}_2/(1-\hat{p}_2)\} \quad . \tag{3.2}$$

The estimate of the variance of f is

$$v_f = \left\{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}\right\} \quad . \tag{3.3}$$

so that an approximate 100(1-α)% confidence interval for O_R can be written as $\exp(f \pm z_{\alpha/2}\sqrt{v_f})$.

Relative risk is a commonly used quantity in epidemiological studies. Relative risk is the risk of developing a particular condition (often a disease) for one group compared to another group. For data collected prospectively, the relative risk is written as $R_R=p_1/p_2$, thus from the definition of odds ratio and relative risk it is true that

$$\hat{R}_R = \hat{O}_R \times \frac{\{1+(n_{21}/n_{22})\}}{\{1+(n_{11}/n_{12})\}} \quad . \tag{3.4}$$

For cross sectional data, the quantity $\hat{p}_1/\hat{p}_2$ does not indicate risk since the disease and risk factor are assessed at the same time. When n_{11} and n_{21} are small relative to n_{12} and n_{22}, that is, the rare outcome assumption holds, $\hat{O}_R$ is approximately equal to $\hat{R}_R$. Usually the outcome of interest needs to occur less than 10% of the time for $\hat{O}_R$ and $\hat{R}_R$ to be similar. For our data we have average asthma rate of 7.81% so it is proper for us to use $\hat{O}_R$ as an approximation to $\hat{R}_R$ in many cases. For details, see Stokes *et al.* (2000, pp. 29-33).

3.2 Application of Odds Ratio Analysis to the NPHS Data

To calculate odds ratio estimates and confidence intervals we just need to combine and adjust variables levels to get 2×2 tables. Thus there will be many odds ratios. The odds ratio results for those we are most interested in appear in Table 3.2. The last column contains a test for H_0: $\hat{O}_R = 1$.

Table 3.2 Results of odds ratio analysis.

Test Meaning	Case-Control (Odds Ratio)	99% CI	χ^2 P-value
Respondents who attend physical activity daily VS not (PAC6DFD)	1.3062	1.2065 1.4141	Pearson 75.4897 <.0001 LLR 73.4328 <.0001
Respondents who attend physical activity ACTIVLLY VS not (PAC6DPAI)	1.3187	1.2074 1.4403	Pearson 65.6666 <.0001 LLR 62.7610 <.0001
Respondents who were born in America & Europe VS other country (SDC6GCB)	0.6511	0.5687 0.7454	Pearson 67.7489 <.0001 LLR 74.5920 <.0001
Respondents who were born in Asia VS other country (SDC6GCB)	0.1957	0.1429 0.2682	Pearson 220.6924 <.0001 LLR 311.3942 <.0001
Respondents who were born in Canada VS other country (SDC6GCB)	2.2713	2.0200 2.5539	Pearson 341.9383 <.0001 LLR 397.9497 <.0001
Respondents who are white VS all others (SDC6GRAC)	1.3356	1.1847 1.5058	Pearson 626.7386 <.0001 LLR 613.3712 <.0001
Respondents who are single VS all others (SDC6GMAR)	1.9125	1.7874 2.0463	Pearson 626.7386 <.0001 LLR 613.3712 <.0001
Respondents who attend physical activity regular VS not (PAC6DFR)	1.1906	1.1005 1.2880	Pearson 32.6571 <.0001 LLR 32.9934 <.0001
The province with the highest asthma rate vs the lowest one (Nova Scotia VS Newfoundland, PRC6_CUR)	1.5062	1.0785 2.1035	Pearson 10.0911 0.0015 LLR 10.4245 0.0012

Test Meaning	Case-Control (Odds Ratio)	99% CI	χ^2 P-value
The province with the asthma rate high than average vs the lower ones (Nova Scotia & Ontario & Quebec VS all others ,PRC6_CUR)	1.0975	1.0217 1.1789	Pearson 11.2069 0.0008 LLR 11.3012 0.0008
The rural AREA vs the urban area (GE36GURB)	0.8166	0.7129 0.9354	Pearson 14.8099 0.0001 LLR 15.1325 0.0001
The Occupation with the highest asthma rate vs the lowest one: (Personal service VS Agriculture, LFC6GI13)	2.5639	1.6839 3.9039	Pearson 35.5754 <.0001 LLR 41.6143 <.0001
The Occupation with the higher than average asthma rate vs the lower ones: (Finance, Community Services & Personal service VS all other occupations, LFC6GI13)	1.5160	1.3692 1.6786	Pearson 111.9838 <.0001 LLR 108.2655 <.0001
Children(age <12) VS Adult (PAC6DEE)	1.6247	1.4929 1.7683	Pearson 221.8114 <.0001 LLR 201.6865 <.0001
Respondents who attend leisure time physical activity VS not (PAC6DLEI)	1.1474	1.0088 1.3051	Pearson 7.5770 0.0059 LLR 7.8128 0.0052
Respondents who are widowed, separated, divorced VS married/common-law/partner (SDC6GMAR)	1.5006	1.3342 1.6878	Pearson 80.314 <.0001 LLR 74.1895 <.0001
Respondents who smoke vs not (SMC6_3)	1.0338	0.9512 1.1236	Pearson 1.0579 0.3037 LLR 1.0520 0.3050
Respondents who start smoke daily before 20 years old vs start after 20 years old (SMC6_3)	1.4853	1.1949 1.8463	Pearson 22.1964 <.0001 LLR 23.8456 <.0001
Respondents who smoke daily more than 10 cig. vs less than 10 (SMC6_4)	0.7644	0.6515 0.8968	Pearson 18.8569 <.0001 LLR 18.8569 <.0001
Respondents who smoke daily more than 10 cig & less than20 years old. vs less than 10 & less than20 years old (SMC6_4)	1.2247	0.8577 1.7486	Pearson 2.1530 0.1423 LLR 2.1461 0.1429
Respondents who smoke daily more than 10 cig & elder than 20 years old. vs less than 10 & elder than 20 years old (SMC6_4)	0.8270	0.6876 0.9946	Pearson 7.0467 0.0079 LLR 6.8573 0.0088
Dwelling not owned by respondents' family member vs yes (DHC6_OWN)	1.4263	1.3276 1.5323	Pearson 164.2046 <.0001 LLR 157.2275 <.0001
Dwelling have no more than 2 bedrooms vs yes (DHC6GBED)	1.1458	1.0633 1.2347	Pearson 22.0446 <.0001 LLR 21.6534 <.0001
FEMALE vs MALE (DHC6_SEX)	1.2036	1.1251 1.2875	Pearson 50.2127 <.0001 LLR 50.2966 <.0001

Test Meaning	Case-Control (Odds Ratio)	99% CI	χ^2 P-value
FOR CHILDREN(<12 YEARS OLD) MALE vs FEMALE (DHC6_SEX)	1.6211	1.3882 1.8931	Pearson 65.2512 <.0001 LLR 65.8816 <.0001
FOR ADULT (AGE>=12 YEARS OLD) FEMALE vs MALE (DHC6_SE)	1.4311	1.3261 1.5445	Pearson 147.9869 <.0001 LLR 148.8596 <.0001
Adult Respondents who are overweight vs not (HWC6GSW)	1.1524	1.0399 1.2771	Pearson 12.6688 0.0004 LLR 12.4570 0.0004
Respondents who have chronic bronchitis or emphysema vs not (CCC6_1H)	6.7796	6.0304 7.6218	Pearson 2316.2080 <.0001 LLR 1391.1395 <.0001

The results in Table 3.2 indicate that the odds ratios significantly depart from 1 in most cases considered. The results go along with the features noted in the exploratory analysis in Section 1.4. For instance, a significantly higher asthma prevalence occurs for respondents born in Canada compared to those born in other countries. A significantly higher asthma prevalence occurs among those who are single compared to those in other marital circumstances. Also a significantly higher asthma prevalence occurs among females when compared to males. On the opposite side, living in rural areas significantly decreases the asthma prevalence when compared to urban areas. Respondents living in dwellings owned by their family members have relatively lower asthma prevalence.

Chapter 4 Loglinear and Logistic Regression Analysis

4.1 Loglinear Regression Models for Two-way s×r Tables and High Dimension Tables

When a sample of n observations is classified with respect to two categorical variables, one having s levels and the other having r levels, then the resulting frequencies can be displayed in an s×r contingency table, as shown below. The corresponding cell probabilities are π_{ij}, with row and column marginal probability π_{i+} and π_{+j}, respectively.

Table 4.1 Representation of an s×r table.

Level of X	Level of Y				Total
	1	2	…	r	
1	n_{11}	n_{12}	…	n_{1r}	n_{1+}
2	n_{21}	n_{22}	…	n_{2r}	n_{2+}
…	…	…	…	…	…
s	n_{s1}	n_{s2}	…	n_{sr}	n_{s+}
Total	n_{+1}	n_{+2}	…	n_{sr}	n

The generalization of the saturated loglinear model for this s×r contingency table is:

$$\log(m_{ij}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}, \qquad i = 1,\dots,s;\ j = 1,\dots,r \ , \tag{4.1}$$

where m_{ij}=n π_{ij} is the expected frequency in the (i,j)th cell. The parameter μ is fixed by the sample size n and the model has s+r+sr parameters. The parameters in the above model satisfy the sum-to zero constraints:

$$\sum_{i=1}^{s}\lambda_i^X = 0, \qquad \sum_{j=1}^{r}\lambda_j^Y = 0 , \qquad \sum_{i=1}^{s}\lambda_{ij}^{XY} = \sum_{j=1}^{r}\lambda_{ij}^{XY} = 0 \ . \tag{4.2}$$

The constraints implies that there only (s-1)+(r-1)+(s-1)(r-1)=sr-1 parameters can be estimated and zero df for testing lack of fit. Letting $\hat{m}_{ij} = n_{i+}n_{+j} / n$, the likelihood ratio statistic

$$G^2 = 2\sum_{i=1}^{s}\sum_{j=1}^{r} n_{ij} \log(n_{ij} / \hat{m}_{ij}) \ . \tag{4.3}$$

test the hypothesis H_0 : $\lambda_{ij}^{XY} = 0$, for i=1,...s-1, j=1,...r-1. Under the null hypothesis of independence, G^2 has an approximate Chi-square distribution with (s-1)(r-1) df. If H_0 is true, the reduced model $\log(m_{ij}) = \mu + \lambda_i^X + \lambda_j^Y$ is the model of independence of X and Y. This model has (s-1)+(r-1) linearly independent parameters of λ and (s-1)(r-1) df for testing lack of fit.

The same logic holds for loglinear models dealing with high dimension tables. For a W way contingency table we can consider all the combination of W-1, ...2 factor interaction terms and all main effects. As the number of dimensions of a contingency table increases, there will be some complicating factors. The number of possible interaction parameters will increase tremendous. This will result in hundreds of or even thousands of parameters to be estimated. Even if the sample size is large, whether these huge number of parameters can be estimated is a problem, since there may be many observed cell counts equal to zero. There may even be marginal totals equal to zero. For details see Stokes *et al.* (2000, pp. 560-568).

In our data set there are 32 variables, some of them having as many as 66 levels. So to construct a loglinear model we need to include only those important variables and combine variable levels to get a reasonable model. Thus we have very a large number of models to chose. All our variables as described in Chapter 1 can be divided into 7 groups. Variables from the same group describe similar or related aspects of respondents. So we chose about one variable from each group. These selected variables all have significant Pearson chi-square and log likelihood ratio chi-square values. Another reason for selecting only one variable from each group is to avoid the multicolinearity problem that arises when highly correlated variables are included. These variables are: CCC6_1C (Has asthma); PRC6_CUR (Province of residence); LFC6GI13 (Industry Codes for main job - 13 groups); PAC6DPAI (Derived physical activity index); SMC6_3 (Age started smoking daily - daily smoker); DHC6_OWN (Dwelling owned by household member); DHC6_SEX (Gender). If we consider all the 6-factor interactions and all its lower order interaction as well as main effects, then we will have thousands of possible parameters to estimate. That is a huge but unnecessary task. We deal with this problem with two methods.

One is to combine the levels of the variables. We combine the l0 levels of province to two: the first level indicating higher asthma prevalence rate, which include Ontario Quebec and PEI. The second level contains all other provinces. We also do a grouping of the 13 levels of occupation. We put those with no occupation in the first level, those occupations with high asthma rate (Finance, community service, personal services, transportation, retail trade, business services, public administration) in the second level and all other occupation in the last level.

The other alternative is to consider hierarchical models. Our aim in constructing loglinear models is to capture the relations among variables. We focus on hierarchy models since non hierarchy models have little meaningful interpretation. By fitting hierarchical models we find the best model which only includes significant interaction terms. We apply the following strategy: we first fit models with all 6-factor interactions, if it is not significant we will try the 5-factor interactions models and keep on till we could confirm that the lower than the present order model which are significant. Then we delete all the terms which are not significant in the present model to get the best one.

4.2 Results of the Seven-way Loglinear Regression Model

The results of the model fitting are listed in the Table 4.2. The 2-factor hierarchy model indicates that not all variables have an effect on CCC6_1C. But the G^2 value is significant, which indicates that we shall consider adding more interaction terms to improve the model fitting. The 5-factor hierarchy model is significant, but the 6-factor one is not. The G^2 value increases from 0.0002 (df=1) for the 6-factor hierarchy model to 65.54 (df =14) for the 5-factor ones. After deleting all the non significant terms in the 6-factor hierarchy model we get the best model which describes the relationship of all the variables. The terms included in the best model and their Chi-square values are listed in table 4.2. Comparing the 5 and 6 factor hierarchy models we see that for the six factor interaction terms only CCC*PRC*LFC*DHC*SMC*DHC6 & CC*LFC*PAC*DHC*SMC*DHC6 are significant and need to be added to the 5-factor hierarchy model to improve the fitting. Many of the 2 and 3 way interactions are significant in the 6 factor hierarchy model. For the three way interaction terms which are related to CCC6_1C, we see that CCC6_1C*PRC6_CU*LFC6GI13, CCC6_1C*PRC6_CU*PAC6DPAI, CCC6_1C*LFC6GI1*PAC6DPAI, CCC6_1C*LFC6GI1*DHC6_SEX, CCC6_*PRC6_*LFC6G*DHC6_S, CCC6_1C*PAC6DPA*DHC6_SEX, CCC6_1C*LFC6GI1*DHC6_OWN, CCC6_1C*PAC6DPA*DHC6_OWN are significant at 5% level.

Table 4.2 Results of loglinear model.

Source	DF	Chi-Square	Pr > ChiSq
CCC6_1C	1	3012.09	<.0001
PRC6_CUR	1	203.27	<.0001
LFC6GI13	2	877.27	<.0001
CCC6_1C*LFC6GI13	2	71.04	<.0001
PRC6_CUR*LFC6GI13	2	31.33	<.0001
CCC6_1C*PRC6_CU*LFC6GI13	2	12.37	0.0021
PAC6DPAI	2	1053.14	<.0001
CCC6_1C*PRC6_CU*PAC6DPAI	2	8.06	0.0178
CCC6_1C*LFC6GI1*PAC6DPAI	2	8.99	0.0112
DHC6_SEX	1	103.7	<.0001
CCC6_1C*DHC6_SEX	1	7.25	0.0071
LFC6GI13*DHC6_SEX	2	406.52	<.0001
CCC6_1C*LFC6GI1*DHC6_SEX	2	21.2	<.0001
PRC6_CU*LFC6GI1*DHC6_SEX	2	8.52	0.0141
CCC6_*PRC6_*LFC6G*DHC6_S	2	6.62	0.0366
PAC6DPAI*DHC6_SEX	2	210.34	<.0001
CCC6_1C*PAC6DPA*DHC6_SEX	2	6.4	0.0408
PRC6_CU*PAC6DPA*DHC6_SEX	2	1.41	0.494
PRC6_*LFC6G*PAC6D*DHC6_S	2	22.11	<.0001
CCC6*PRC6*LFC6*PAC6*DHC6	2	22.95	<.0001
SMC6_3	1	1016.43	<.0001
PRC6_CUR*SMC6_3	1	4.27	0.0388
LFC6GI13*SMC6_3	2	59.17	<.0001
PRC6_CUR*LFC6GI13*SMC6_3	2	9.55	0.0084
CCC6_*PRC6_*LFC6G*SMC6_3	2	10.97	0.0042
PAC6DPAI*SMC6_3	1	7.64	0.0057
CCC6*PRC6*LFC6*PAC6*SMC6	2	7.55	0.023
CCC6_1C*DHC6_SEX*SMC6_3	1	25.26	<.0001
PRC6_CUR*DHC6_SEX*SMC6_3	1	18.83	<.0001
CCC6_*PRC6_*DHC6_*SMC6_3	1	24.96	<.0001
LFC6GI13*DHC6_SEX*SMC6_3	2	18.99	<.0001
CCC6_*LFC6G*DHC6_*SMC6_3	2	19.31	<.0001
CCC6*PRC6*LFC6*DHC6*SMC6	2	10.2	0.0061
CCC6_*PAC6D*DHC6_*SMC6_3	1	4.46	0.0347
LFC6G*PAC6D*DHC6_*SMC6_3	2	8.46	0.0145
CCC6*LFC6*PAC6*DHC6*SMC6	2	7.44	0.0242
DHC6_OWN	1	435.56	<.0001
PRC6_CUR*DHC6_OWN	1	4.34	0.0371
LFC6GI13*DHC6_OWN	2	18.41	0.0001
CCC6_1C*LFC6GI1*DHC6_OWN	2	6.48	0.0392
PRC6_CU*LFC6GI1*DHC6_OWN	2	9.33	0.0094
CCC6_*PRC6_*LFC6G*DHC6_O	2	2.04	0.36
PAC6DPAI*DHC6_OWN	2	30.68	<.0001
CCC6_1C*PAC6DPA*DHC6_OWN	2	12.28	0.0022
DHC6_SEX*DHC6_OWN	1	4.09	0.043

LFC6GI1*DHC6_SE*DHC6_OWN	2	6.4	0.0408
CCC6_*LFC6G*DHC6_*DHC6_O	2	7.12	0.0284
PRC6_*LFC6G*DHC6_*DHC6_O	2	6.54	0.038
LFC6G*PAC6D*DHC6_*DHC6_O	2	18.19	0.0001
CCC6*LFC6*PAC6*DHC6*DHC6	2	19.12	<.0001
SMC6_3*DHC6_OWN	1	152.52	<.0001
CCC6_*LFC6G*SMC6_*DHC6_O	2	9.88	0.0072
PRC6_*LFC6G*SMC6_*DHC6_O	2	0.17	0.92
PAC6DPAI*SMC6_3*DHC6_OWN	1	11.81	0.0006
CCC6_*PAC6D*SMC6_*DHC6_O	1	4.66	0.0309
PRC6_*PAC6D*SMC6_*DHC6_O	1	6.18	0.0129
LFC6G*PAC6D*SMC6_*DHC6_O	2	9.62	0.0081
DHC6_SEX*SMC6_3*DHC6_OWN	1	6.02	0.0141
CCC6_*DHC6_*SMC6_*DHC6_O	1	4.57	0.0326
PRC6_*DHC6_*SMC6_*DHC6_O	1	4.37	0.0365
CCC6*PRC6*DHC6*SMC6*DHC6	1	2.63	0.105
LFC6G*DHC6_*SMC6_*DHC6_O	2	14.1	0.0009
CCC6*LFC6*DHC6*SMC6*DHC6	2	16.81	0.0002
PRC6*LFC6*DHC6*SMC6*DHC6	2	12.41	0.002
CCC*PRC*LFC*DHC*SMC*DHC6	2	9.76	0.0076
LFC6*PAC6*DHC6*SMC6*DHC6	2	29.22	<.0001
CCC*LFC*PAC*DHC*SMC*DHC6	2	41.26	<.0001

4.3 Logistic, Probit & loglog Regression Models

Suppose we have N binary responses as independent Bernoulli radom variables. Let $x_i=(x_{i0}, x_{i1},\ldots, x_{ik})$ denote the ith setting of values of k explanatory variables, i=1,2,...I, where $x_{i0}=1$. when more than one observation on Y occurs at a fixed x_i value, it is sufficient to record the number of observations n_i and the number of "1" outcomes. Thus let Y_i refer to this "success" count rather than to individual binary responses. The { Y_i i=1,...,I} are independent binomial random variables with $E(Y_i)=n_i\pi(x_i)$, where $N=n_1+\ldots+n_I$. The joint probability mass function of $(Y_1,\ldots, Y_I)$ is proportional to the product of I binomial terms, so we have:

$$\prod_{i=1}^{I} \pi(x_i)^{y_i}[1-\pi(x_i)]^{n_i-y_i} = \left\{\prod_{i=1}^{I}[1-\pi(x_i)]^{n_i}\right\}\left\{\prod_{i=1}^{I}\exp\left[\log\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right)^{y_i}\right]\right\}$$

$$= \left\{\prod_{i=1}^{I}[1-\pi(x_i)]^{n_i}\right\}\exp\left[\sum y_i \log\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right)\right] \quad . \tag{4.4}$$

Let $\pi(x) = \frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)}$, after transformation we get $\log\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = \eta = \sum_j \beta_i x_{ij}$,

after some rearrangement we can write the log likelihood function as:

$$L(\beta) = \sum_j\left(\sum_i y_i x_{ij}\right)\beta_j - \sum_i n_i \log\left[1+\exp\left(\sum_j \beta_j x_{ij}\right)\right] \quad . \tag{4.5}$$

this depends on the binomial counts only through the sufficient statistics $\sum_i y_i x_{ij}$, $j=0,\ldots,k$. By using a numeric procedure such as Newton-Raphson method we can get the ML estimate of β_i, thus we can get an estimate of $\pi(x_i)$.

Alternatively define $\pi(x) = \Phi[(x-\mu)/\sigma]$, where Φ is the standard normal cdf. If we use $\Phi^{-1}[\pi(x)] = \alpha+\beta x$ the model is the probit model. If we define $\pi(x) = \exp[-\exp(\alpha+\beta x)]$ and use $\log\{-\log[\pi(x)]\} = \alpha+\beta x$, we are fitting a complementary log-log model. All these three types of models are considered in our analysis. For details see Agresti (1990, pp. 102-115).

Our data set contains various kinds of variables. Categorical and continuous, ordinal and non ordinal all exist. There are also many structural missing values, such as variable SMC_3, which only indicates the starting age of respondents who smoke. Thus before we can fit the logistic regression model we need to rearrange the data.

For the categorical and non ordinal variables, the code values just indicate difference categories. It is not proper to put them directly to the design matrix of x. We need to create dummy variables for them. To fit our logistic and other two kinds of regression models, we use the reference code. That is, we select a base category, others effects are all compared to this base one. Parameter estimates using the reference coding scheme estimate the difference in the effect of each non-reference level compared to the effect of the reference level.

For ordinal variables we first look whether the increase of code value respondent to uniform change of the object. For example, the grouped variable DHC6GAGE. The variable describe the respondent's age. But we only get the grouped value due to confidentiality requirements. The one unit increase of code corresponds to an increase of 5 years. Thus for this variable we include it in the models using its own code value. Fortunately all the categorical ordinal variables are like this. Continuous variables such as PAC6DFM (monthly freq. of physical activity) are all included in the models directly.

It is a little complicated to treat the structural missing values. For example, variable LFC6GI13 describes the respondent's occupation. For those who do not work there will be at this variable if we include LFC6GI13 in the model. SAS will automatically delete cases with missing values when running a logistic regression procedure. Since we have many variables with structural missing values, if we do not rearrange the data then only a small number of cases can be left for us to do analysis. So for this variable we treat the respondents who do not have occupation as another category and make a dummy variable for it except the 12 dummy variables respondent to its levels. SMC6_3 (age start smoking) is another categorical variable with structural missing value due to respondents who do not smoke. If viewing those persons who do not smoke as they start smoking very late, say, in their 100 years old, then we could eliminate the structural missing value by substituting them with a large number say, 100. All the details about the rearranged codes and information about the design matrix are listed in Appendix A.

4.4 Results of the Logistic, Probit, and Complementary Log-log Regression Models for All Respondents

The results in Tables 4.3-4.5 were obtained by using stepwise regression method in SAS. In addition to the log likelihood ratio test, parameters estimate, odds ratio estimates and some other statistics, we also do the Hosmer and Lemeshow goodness-of-fit test. Receiving operating characteristic curve (ROC) is also plotted using Splus.

Table 4.3 Result of logistic regression.

```
Step 37. Effect DHC6DL5 entered:
                      Model Convergence Status
           Convergence criterion (GCONV=1E-8) satisfied.

                        Model Fit Statistics
                                            Intercept
                               Intercept       and
               Criterion         Only       Covariates
               AIC            39793.480      38201.469
               SC             39802.626      38512.448
               -2 Log L       39791.480      38133.469

               Testing Global Null Hypothesis: BETA=0
       Test                 Chi-Square       DF     Pr > ChiSq
       Likelihood Ratio      1658.0108       33         <.0001
       Score                 1547.9753       33         <.0001
       Wald                  1440.1822       33         <.0001

                      Residual Chi-Square Test
                  Chi-Square        DF     Pr > ChiSq
                    109.7621        32         <.0001
NOTE: No (additional) effects met the 0.05 significance level for entry into the model.
```

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-3.8002	0.1798	446.8662	<.0001
NFLD	1	-0.3610	0.1140	10.0250	0.0015
NB	1	-0.2023	0.0937	4.6595	0.0309
AB	1	-0.1492	0.0500	8.8937	0.0029
GE36GU1	1	-0.1730	0.0503	11.8512	0.0006
HWC6G1	1	-0.2285	0.0701	10.6107	0.0011
HWC6G2	1	-0.3293	0.0404	66.4116	<.0001
HWC6G3	1	-0.1705	0.0527	10.4460	0.0012
LFC1	1	0.4606	0.0493	87.3481	<.0001
LFC3	1	0.2319	0.0881	6.9243	0.0085
LFC6	1	0.2472	0.0854	8.3843	0.0038
LFC9	1	0.6246	0.0877	50.6819	<.0001
LFC10	1	0.4770	0.0600	63.2354	<.0001
LFC11	1	0.3843	0.0682	31.7653	<.0001
LFC12	1	0.2770	0.0740	14.0131	0.0002
LFC13	1	0.1982	0.0938	4.4629	0.0346
PAC6DPA1	1	0.5385	0.0586	84.5208	<.0001
PAC6DPA2	1	0.5149	0.0618	69.5356	<.0001
PAC6DPA3	1	0.4139	0.0602	47.2899	<.0001
SDC6G1	1	1.0289	0.1382	55.4054	<.0001
SDC6G2	1	0.8823	0.1477	35.7008	<.0001
SDC6G3	1	-0.7150	0.1943	13.5479	0.0002
SMC6_31	1	-0.00140	0.000442	10.0113	0.0016

DHC6_O2	1	0.3001	0.0333	81.2565	<.0001
DHC6_S2	1	0.1129	0.0299	14.2879	0.0002
DHC6GM2	1	0.3870	0.0535	52.3058	<.0001
DHC6GM3	1	0.5043	0.0665	57.5186	<.0001
DHC6GE1	1	0.1448	0.0499	8.4107	0.0037
DHC6GE4	1	0.3022	0.0808	13.9754	0.0002
DHC6GE5	1	0.3616	0.0692	27.3423	<.0001
DHC6GE6	1	0.3671	0.0614	35.7731	<.0001
DHC6DL5	1	0.1592	0.0784	4.1234	0.0423
DHC6DL6	1	0.3509	0.0844	17.2851	<.0001
DHC6GAGE	1	-0.0718	0.00662	117.5309	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
NFLD	0.697	0.557	0.871
NB	0.817	0.680	0.982
AB	0.861	0.781	0.950
GE36GU1	0.841	0.762	0.928
HWC6G1	0.796	0.694	0.913
HWC6G2	0.719	0.665	0.779
HWC6G3	0.843	0.760	0.935
LFC1	1.585	1.439	1.746
LFC3	1.261	1.061	1.499
LFC6	1.280	1.083	1.514
LFC9	1.867	1.572	2.218
LFC10	1.611	1.433	1.812
LFC11	1.469	1.285	1.679
LFC12	1.319	1.141	1.525
LFC13	1.219	1.014	1.465
PAC6DPA1	1.713	1.528	1.922
PAC6DPA2	1.674	1.483	1.889
PAC6DPA3	1.513	1.344	1.702
SDC6G1	2.798	2.134	3.669
SDC6G2	2.417	1.809	3.228
SDC6G3	0.489	0.334	0.716
DHC6_O2	1.350	1.265	1.441
DHC6_S2	1.119	1.056	1.187
DHC6GM2	1.473	1.326	1.635
DHC6GM3	1.656	1.453	1.886
DHC6GE1	1.156	1.048	1.275
DHC6GE4	1.353	1.155	1.585
DHC6GE5	1.436	1.254	1.644
DHC6GE6	1.443	1.280	1.628
DHC6DL5	1.173	1.006	1.367
DHC6DL6	1.420	1.204	1.676

Association of Predicted Probabilities and Observed Responses

Percent Concordant	60.5	Somers' D	0.227
Percent Discordant	37.8	Gamma	0.231
Percent Tied	1.6	Tau-a	0.032
Pairs	337650512	c	0.613

Partition for the Hosmer and Lemeshow Test

Group	Total	CCC6_1C = YES Observed	CCC6_1C = YES Expected	CCC6_1C = NO Observed	CCC6_1C = NO Expected
1	6946	251	188.78	6695	6757.22
2	6934	350	296.75	6584	6637.25
3	6861	375	347.82	6486	6513.18
4	6907	416	394.89	6491	6512.11
5	6939	443	443.73	6496	6495.27
6	6916	458	500.72	6458	6415.28
7	6932	625	579.97	6307	6352.03
8	6932	715	683.64	6217	6248.36
9	6962	740	820.45	6222	6141.55
10	6989	899	1082.47	6090	5906.53

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
89.5752	8	<.0001

Table 4.4 Result for probit regression.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	39793.480	38213.265
SC	39802.626	38515.098
-2 Log L	39791.480	38147.265

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1644.2145	32	<.0001
Score	1542.7256	32	<.0001
Wald	1460.9200	32	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
102.2374	31	<.0001

NOTE: No (additional) effects met the 0.05 significance level for entry into the model.

Association of Predicted Probabilities and Observed Responses

Percent Concordant	60.5	Somers' D	0.227
Percent Discordant	37.9	Gamma	0.230
Percent Tied	1.6	Tau-a	0.032
Pairs	337650512	c	0.613

Partition for the Hosmer and Lemeshow Test

Group	Total	CCC6_1C = YES Observed	CCC6_1C = YES Expected	CCC6_1C = NO Observed	CCC6_1C = NO Expected
1	6964	246	183.41	6718	6780.59
2	6918	353	293.42	6565	6624.58
3	6917	377	352.17	6540	6564.83
4	6956	413	403.66	6543	6552.34
5	6906	454	450.18	6452	6455.82
6	6939	458	513.60	6481	6425.40
7	6926	623	592.18	6303	6333.82
8	6925	713	692.05	6212	6232.95
9	6931	755	819.28	6176	6111.72
10	6936	880	1053.74	6056	5882.26

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
85.1368	8	<.0001

Table 4.5 Result for Complementary log-log.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	39793.480	38199.962
SC	39802.626	38510.942
-2 Log L	39791.480	38131.962

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1659.5177	33	<.0001
Score	1547.9753	33	<.0001
Wald	1459.6394	33	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
102.8374	30	<.0001

NOTE: No (additional) effects met the 0.05 significance level for entry into the model.

Association of Predicted Probabilities and Observed Responses

Percent Concordant	60.5	Somers' D	0.227
Percent Discordant	37.8	Gamma	0.231
Percent Tied	1.6	Tau-a	0.032
Pairs	337650512	c	0.613

Partition for the Hosmer and Lemeshow Test

		CCC6_1C = YES		CCC6_1C = NO	
Group	Total	Observed	Expected	Observed	Expected
1	6918	250	189.54	6668	6728.46
2	6901	348	296.24	6553	6604.76
3	6870	378	348.22	6492	6521.78
4	6902	420	393.83	6482	6508.17
5	6942	436	442.41	6506	6499.59
6	6917	455	498.63	6462	6418.37
7	6939	628	577.76	6311	6361.24
8	6922	715	679.45	6207	6242.55
9	6945	736	814.87	6209	6130.13
10	7062	906	1096.52	6156	5965.48

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
92.6827	8	<.0001

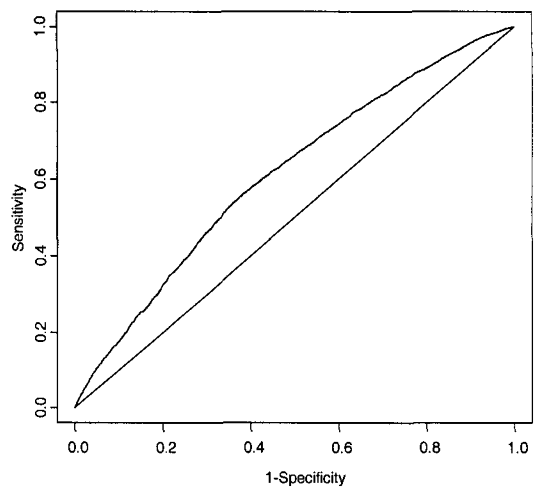


Figure 4.1 Receiving operating characteristic curve (ROC).

From the above results (refer to the programs and the variable meaning listed in the appendix A and H) we see that for all three types of regression the results are very similar. The likelihood ratio chi-square statistic are all large enough to justify the goodness of the model. But for all three models the Hosmer and Lemeshow goodness-of-fit Test shows some lack of fit in the models. Generally say, the variables included in the models are important and statistically significant, but there may still exist some other important factors to be included in the models. Consider the fact that our data set only contains the environmental factors thought to effect asthma prevalence but there could exist other important variables such as the genetic ones. It may be those variables that cause Hosmer and Lemeshow value in these models significant. In the next section we will apply an approach to alleviate this problem considering in fitting separate regression for infants, children and adults. We will also address problem of model validation.

To assess the ability of the logistic model to discriminate between events and non-events the receiver operating characteristic (ROC) curve is a good diagnostic tool. For every respondent in the data set we can get the corresponding estimated probability of event. Sort these estimated probabilities and view every one of them as cut-points for predicting the response, then label any observation with an estimated event probability that exceeds or equals a given threshold (or cut-points) as being predicted to be an event. Calculate the proportion of event observations that were correctly predicted to have an event response, this is called the sensitivity corresponding to the cut-points. We can also get the relevant one minus specificity by calculating the proportion of nonevent observations that were predicted to have an event response. ROC is the curve obtained by plotting sensitivity against one minus specificity. The ROC curve passes through the points (0, 0) and (1, 1) and lift up right. The area below the curve is one when the model discriminates perfectly, whereas the ROC curve lines from (0, 0) to (1, 1) when the model has no discriminating ability. In SAS the area below the ROC curve is given out and labeled as "c" under the title "Association of Predicted Probabilities and Observed

Responses". From the plot of our models ROC curve we see that the curve is moderately satisfying and the statistic c is calculated as 0.613. This indicates that our models are reasonably useful and asthma prevalence rate can be explained by variation of environmental factors to some extent.

From the maximum likelihood analysis we see that for the ten provinces only the influence of Newfoundland, New Brunswick and Alberta are significant. These three provinces all have asthma prevalence lower than the average asthma prevalence and the negative value of their parameters indicates that living in these three provinces have a positive influence on people's health. People living in rural areas (GE36GU1) have a lower asthma prevalence. Adult with acceptable body mass (HWC6G2) tend to have less asthma prevalence than those with insufficient or some excess body mass. The parameters related to occupational variables indicate that the occupations of manufacturing non durable goods (LFC3), transportation (LFC6), finance (LFC9), community services (LFC10), personal services (LFC11), business services (LFC12) and public administration (LFC13) have a significantly adverse effect on people's health concerning asthma. People who engage in these occupations have a higher asthma prevalence. The estimate value of PAC6DPA1, PAC6DPA1 and PAC6DPA3 convey the relationship of physical activities with asthma. People who have less physical activities tend to get less asthma. This is consistent with our previous results and some findings in the asthma studies. Though physical activities do good to people, they also may increase the asthma prevalence. Relative research on asthma also indicates that there are many cases of exercise-induced asthma (see Manning, 1993, Tan *et al.* 2002 and website on Asthma and Allergy Information and Research (AAIR)). People born in Asia have much less asthma prevalence according to the large negative value of SDC6G3, while people born in Canada have the highest asthma prevalence over those born in Asia, Europe and USA. People who start smoking early (SMC6_31) have a higher asthma prevalence. The results also indicate that people who do not own the dwellings they live in (DHC6_O2), who are female (DHC6_S2) or who are young (DHC6GAGE) have high asthma prevalence. The asthma prevalence for people who are married or have common-law or partner is smaller than those who are single (DHC6GMA2), widowed, separated or divorced (DHC6GMA3). Parameter estimates for variables representing household type and living arrangement (DHC6GE1, DHC6GE 4, DHC6GE5, DHC6GE 6, DHC6DL5, DHC6DL6) all show that those households in which parents are living with children have higher asthma prevalence.

From the results of the logistic regression we conclude that environmental factors do have a significant influence on asthma prevalence. It can to some extent explain the asthma distribution in the population.

4.5 Logistic Regression Analysis for Infants, Children and Adults

A possible difficulty with regression models fitted to the entire data set is that the relationships between response and predictors assumed by the models vary with age. In an attempt to examine this problem, we will fit separate regression models to the following age groups: infants (<=5 years), children (6-19 years) and adults (>=20 years). The sizes of these groups appear in Table 4.6. Note that the sizes of each group do not included observations with missing values.

Table 4.6 Age-group sizes.

	Infants	Children	Adults	total
Number of subjects	3,107	9,675	56,536	69,318

The logistic regression model was fitted to each age group following the approach described in Section 4.4 for the entire data, the results are reported in Table 4.7. ROC curves were also plotted in Figure 4.2.

Table 4.7 Stepwise logistic regression for infants, children and adults.

- Stepwise Logistic Regression Result for Infants.

```
                 Model Convergence Status
       Convergence criterion (GCONV=1E-8) satisfied.

                   Model Fit Statistics
                                          Intercept
                            Intercept        and
            Criterion         Only       Covariates
            AIC             2576.873       2367.838
            SC              2582.914       2428.252
            -2 Log L        2574.873       2347.838

            Testing Global Null Hypothesis: BETA=0
    Test                  Chi-Square       DF     Pr > ChiSq
    Likelihood Ratio        227.0348        9         <.0001
    Score                   236.3920        9         <.0001
    Wald                    198.1016        9         <.0001

                  Residual Chi-Square Test
              Chi-Square       DF     Pr > ChiSq
                 30.0237       21         0.0915

           Analysis of Maximum Likelihood Estimates
                                   Standard
Parameter     DF     Estimate        Error     Chi-Square     Pr > ChiSq
Intercept      1      -3.8072       0.3347       129.4231         <.0001
SASK           1      -1.8112       0.7219         6.2944         0.0121
DHC6_O2        1       0.6451       0.1276        25.5508         <.0001
DHC6_S2        1      -0.9299       0.1174        62.6915         <.0001
DHC6GE5        1       1.1654       0.2451        22.6065         <.0001
DHC6GE9        1       4.5139       0.8729        26.7407         <.0001
DHC6DL6        1       1.6858       0.2248        56.2591         <.0001
DHC6DL9        1       0.9387       0.1711        30.0996         <.0001
HWC6GHT        1       0.0722       0.0112        41.7066         <.0001
DHC6GAGE       1      -0.8514       0.1378        38.1976         <.0001

   Association of Predicted Probabilities and Observed Responses
        Percent Concordant         63.0    Somers' D    0.279
        Percent Discordant         35.1    Gamma        0.284
        Percent Tied                1.8    Tau-a        0.034
        Pairs                    581400    c            0.640

           Hosmer and Lemeshow Goodness-of-Fit Test
              Chi-Square         DF      Pr > ChiSq
                 49.4026          8          <.0001
```

- Stepwise Logistic Regression Result for Children

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	11392.583	10902.297
SC	11399.760	11081.730
-2 Log L	11390.583	10852.297

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	538.2855	24	<.0001
Score	530.4603	24	<.0001
Wald	408.0939	24	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
68.1231	33	0.0003

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-4.5118	0.4270	111.6541	<.0001
QUE	1	0.1511	0.0612	6.1009	0.0135
BC	1	0.2641	0.0746	12.5286	0.0004
GE36GU1	1	-0.2246	0.0833	7.2652	0.0070
LFC3	1	-0.6772	0.3038	4.9697	0.0258
LFC4	1	-1.3616	0.4711	8.3541	0.0038
LFC5	1	-0.7868	0.2942	7.1534	0.0075
LFC6	1	1.3229	0.2690	24.1775	<.0001
LFC7	1	0.8335	0.2450	11.5792	0.0007
LFC8	1	-0.3738	0.1308	8.1707	0.0043
LFC9	1	2.2393	0.2694	69.1067	<.0001
PAC6DPA2	1	0.3252	0.0712	20.8345	<.0001
SDC6G1	1	3.0832	0.3845	64.3091	<.0001
SDC6G2	1	1.9178	0.4723	16.4867	<.0001
DHC6_O2	1	0.3432	0.0645	28.3153	<.0001
DHC6GB1	1	0.0845	0.0329	6.5891	0.0103
DHC6_S2	1	-0.1402	0.0514	7.4297	0.0064
DHC6GE3	1	0.9830	0.2366	17.2568	<.0001
DHC6GE6	1	1.4646	0.4027	13.2305	0.0003
DHC6DL2	1	0.6553	0.2394	7.4907	0.0062
DHC6DL5	1	1.6181	0.4924	10.8005	0.0010
DHC6DL6	1	0.2610	0.1178	4.9120	0.0267
DHC6DL8	1	0.3558	0.0864	16.9744	<.0001
HWC6GHT	1	-0.0399	0.00572	48.7487	<.0001
DHC6GAGE	1	0.2536	0.0426	35.3919	<.0001

Association of Predicted Probabilities and Observed Responses

Percent Concordant	55.0	Somers' D	0.115
Percent Discordant	43.5	Gamma	0.117
Percent Tied	1.5	Tau-a	0.025
Pairs	10126314	c	0.557

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
92.4581	8	<.0001

- Stepwise Logistic Regression Result for Adults.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	25049.257	24257.856
SC	25058.200	24481.422
-2 Log L	25047.257	24207.856

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	839.4011	24	<.0001
Score	797.5723	24	<.0001
Wald	763.6152	24	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
114.9018	38	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-0.9860	0.4695	4.4101	0.0357
NFLD	1	-0.6575	0.1705	14.8646	0.0001
NB	1	-0.3329	0.1271	6.8622	0.0088
QUE	1	-0.1011	0.0485	4.3480	0.0371
ONT	1	0.1385	0.0433	10.2546	0.0014
HWC6GB1	1	0.00921	0.00226	16.6009	<.0001
HWC6G2	1	-0.1969	0.0399	24.3546	<.0001
LFC1	1	0.3722	0.0493	57.0487	<.0001
LFC2	1	-0.4668	0.1915	5.9446	0.0148
LFC7	1	-0.4472	0.1403	10.1591	0.0014
LFC9	1	0.2247	0.0914	6.0391	0.0140
LFC10	1	0.2411	0.0550	19.2338	<.0001
PAC6DPA1	1	0.1412	0.0452	9.7538	0.0018
SDC6G1	1	0.8223	0.1411	33.9749	<.0001
SDC6G2	1	0.7091	0.1502	22.2745	<.0001
SDC6G3	1	-0.7501	0.2015	13.8527	0.0002
DHC6_O2	1	0.2554	0.0455	31.5535	<.0001
DHC6GB1	1	-0.0506	0.0210	5.8018	0.0160
DHC6_S2	1	0.1652	0.0543	9.2641	0.0023
DHC6GM3	1	0.2135	0.0529	16.2800	<.0001
DHC6GE4	1	0.2751	0.0836	10.8310	0.0010
DHC6GE5	1	0.2647	0.0819	10.4438	0.0012
DHC6DL4	1	-0.2465	0.0443	30.8918	<.0001
HWC6GHT	1	-0.0337	0.00692	23.7150	<.0001
DHC6GAGE	1	-0.0776	0.00802	93.5061	<.0001

Association of Predicted Probabilities and Observed Responses

Percent Concordant	60.4	Somers' D	0.227
Percent Discordant	37.8	Gamma	0.231
Percent Tied	1.8	Tau-a	0.029
Pairs	204207724	c	0.613

Partition for the Hosmer and Lemeshow Test

Group	Total	CCC6_1C = YES Observed	CCC6_1C = YES Expected	CCC6_1C = NO Observed	CCC6_1C = NO Expected
1	5622	211	141.37	5411	5480.63
2	5688	262	219.57	5426	5468.43
3	5623	267	258.78	5356	5364.22
4	5619	305	296.30	5314	5322.70
5	5642	341	335.99	5301	5306.01
6	5661	340	377.48	5321	5283.52
7	5650	404	421.88	5246	5228.12
8	5668	440	480.04	5228	5187.96
9	5653	544	559.92	5109	5093.08
10	5710	764	778.68	4946	4931.32

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
53.6058	8	<.0001

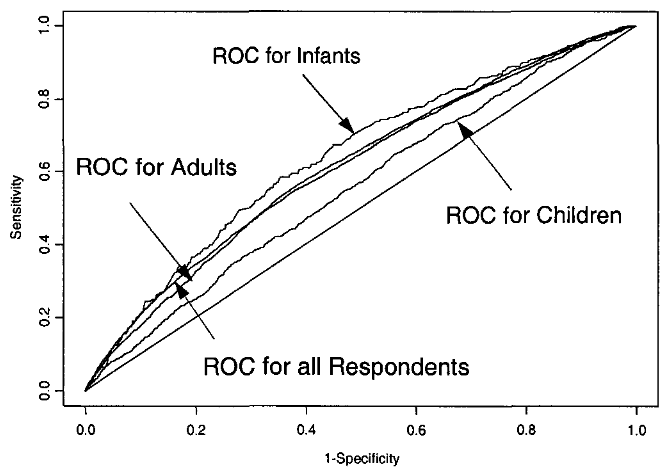


Figure 4.2 ROC curve for infants, children, adults and all respondents

From the results we see that the log likelihood ratio chi-square statistics for stepwise logistic regression models of infants, children and adults are all large enough to justify the significance of the model. However, all three models still show some lack of fit by the Hosmer and Lemeshow tests. Stepwise logistic regression results of all three groups of respondents show that there are relations between asthma prevalence and outdoor environment (provinces they live in), indoor environment (whether the dwelling they living is owned by family member, their living arrangement, etc.), and their basic physical characteristics (gender, height, age, etc.). The results for adults are similar to the results we get in section 4.4 for all respondents. However, they differ from those for infants and children. This may be explained by differences in the relationships across the groups but also by the fact that the adults group is the largest and thus this age group commands the trend in the entire data set. As would be expected, the ROC plot shown in Figure 4.2 shows that the ROC curve for adults and all respondents look very similar. The ROC curve for infants is relatively higher and the ROC curve for children is very low. These results indicate that for infants and adults environmental factors contained in our data set have moderate predictive power on asthma prevalence, but for children the predictive power is relatively low. This is an interesting point that deserves further research. We suspect that this may be caused by the fact that most children will spend much of their time in schools. Leickly (2003) estimated that people spend anywhere from 30 to 60% of their time in homes with the majority of the remaining time in an enclosed work area, or if it is a child, an indoor school environment. So other variables that describe school environment could be important to the study of the relation between children's asthma prevalence and environment.

4.6 Logistic Regression Model Validation

To acess the reliability of the logistic regression model, some limited cross validation was conducted. The subjects within each age group were randomly split into three subgroups using a trinomial distribution with equal probabilities (1/3, 1/3, 1/3). The first subgroup (Part A) was used for variable selection through stepwise logistic regression. The second subgroup (Part B) was used to fit the logistic regression model with the variables chosen in Part A. the third subgroup (Part C) is used to test the model. Figure 4.3 displays the ROC curves obtained from the test data (Part C) along with the ROCs for the groups from Figure 4.2.

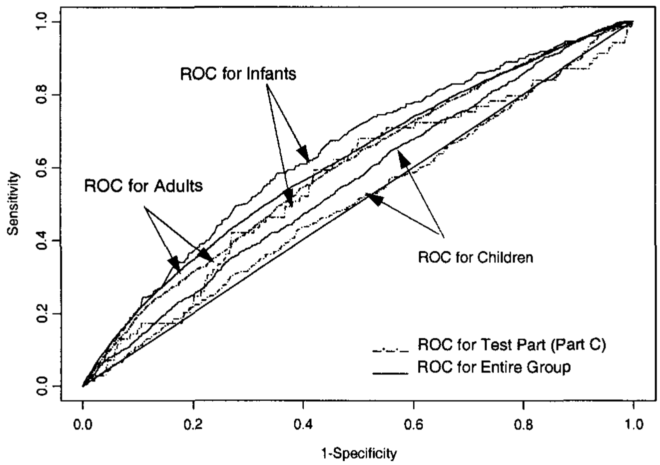


Figure 4.3 Cross validation ROC curve for infants, children and adults.

ROC curves of adults for entire group and test part are very similar to each other. This indicates the reliability of the results for adults. The corresponding ROC curves for infants and children in the entire groups and test parts differ from each other. However the ROC curves of infants are still high. Roc curves for children are all low, especially the ROC for the test part. From these results we conclude that the model fit results for infants and adults are reliable but that is not the case for children. Other factors may be need to study children's asthma prevalence as far as the logistic regression approach is concerned.

Chapter 5: Conclusions

The studies show that asthma rates are significantly different among provinces. The provinces of Nova Scotia and Newfoundland have the highest and lowest asthma rates, respectively. Compared to people living in urban areas, people living in rural areas have a significantly lower asthma prevalence. The asthma rate is also significantly different among the 26 or 33 regions divided by health area of the three provinces Ontario, Manitoba and Alberta. The overall message is that the regional outdoor environment influences the asthma prevalence rate.

Working environment is another factor that influences asthma prevalence and relative risk. Compared to people who have an occupation those who do not work have a higher asthma rate. For people who engage in different occupation the asthma risk they face varies significantly. People engaged in agriculture have the smallest asthma rate among all occupations. No matter whether we use the 21 occupations or the 13 occupations generated by the different classifications, people who engage in occupations related to service have relative higher asthma prevalence. Our study shows that finance, community services have an adverse effect on asthma.

Physical exercise influences the air change amount of our respiratory system. Iatrical studies find so many cases with exercise-induced asthma. Our studies also show that there is some relationship between the amount of physical exercise and asthma. Studies of our data which come from the NPHS survey of Canada suggest that people who have more physical exercise tend to have relatively higher asthma prevalence than others. Many cases with exercise-induced asthma have their first asthma attack during heavy exercise. It is deemed that the large amount and sudden inhale of cold and dry air trigger the symptoms of asthma. Warm up before exercise can lessen the stimuli to airway. Recent paper of Tan and Spector (2002) also indicate the benefits of warm up. So some exercise induced asthma can be prevented or lessen the asthma symptoms by not exercising so much and warming up sufficiently before exercising.

Smoking is another factor we consider in this study. Our study results show that for those who smoke the age people start smoking has very significant influence on their asthma prevalence. The earlier they start smoking, the higher the asthma prevalence among them. But among smokers the number of cigarettes they consume daily has no significant influence on asthma prevalence.

Our results show that demographic and social factors are another aspect we should consider when dealing with asthma. People who were born in Canada, America and Europe and Asia have asthma prevalence from the highest to the lowest one. Whether one has a well-balanced family structure can also have a significant influence on asthma rate. People who are married or have common-law or partner have a lower asthma prevalence compare to those who are single, widowed, separated or divorced.

Indoor environment can be another factor influencing asthma. Our study shows that the number of bedrooms in the dwelling, whether the dwelling is owned by household member, the household type and living arrangement all have an influence on asthma prevalence. Those people who live in dwellings owned by their family members, the dwelling they living in have more than two bedrooms have a less asthma prevalence.

The study of the respondents' basic characteristics shows that gender is an influential factor for asthma. From the results of our analysis, we found that females have higher asthma prevalence than males. Children normally have higher asthma prevalence than adults. Adults with weight in the middle range have a lower asthma prevalence than those who are underweight or overweight.

Considering the above results we see that regional outdoor environment, working environment, physical exercise and air exchange amount, demographic and social environment, indoor environment and smoking pattern all affect asthma prevalence. They can to some extend explain the asthma distribution in the population.

These findings are consistent with results reported by other researchers, including Lierl and Hornung (2003), Maffei *et al.* (2001), FitzGerald *et al.* (2001), Ree and Kanabar (2000) and Tan and Spector (2002). Note that the data we used in this project is a cross sectional one. That is, the data are not collected prospectively. Thus some confounding may occur due to the fact that asthma may make people change their environment and behavior.

Our study is based on the Health File of the Public Use Microdata Documentation from the 1996-1997 National Population Health Survey. The Health File includes household residents in all provinces with the exclusion of populations on Indian Reserves, Canadian Forces Bases and some remote areas in Québec and Ontario. Accordingly, the findings in this project are applicable only to the scope of the Health File.

REFERENCES

1. Alan, A. (1990), *Categorical Data Analysis*, New York: John Wiley & sons, Inc. pp. 44, 102-115.
2. Bruce, K. R., Michael T. N., and Peter, J. B. (1998), *Conquering Childhood Asthma: An Illustrated Guide to Understanding the Treatment and Control of Childhood Asthma*, Hamilton: B.C. Decker, Inc.
3. Chap. T. L. (1998), *Applied Categorical Data Analysis*, New York: John Wiley & sons, Inc.
4. FitzGerald, M., Ernst, P., Boulet, L. P., and O'Byrne, M. P. (2001), *Evidence-based Asthma Management*, Hamilton: B.C. Decker.
5. Leickly, E. F. (2003), "Children, their School Environment, and Asthma", *Annals of Allergy, Asthma, and Immunology*, 90: 3-6.
6. Lierl, B. M., and Hornung, W. R. (2003), "Relationship of Outdoor Air Quality to Pediatric Asthma Exacerbations", *Annals of Allergy, Asthma, and Immunology*, 90:28-34.
7. Maffei, V. J., Beckett, S. W., Belanger, K., Triche, E., Zhang, H., Machung, F. J., and Leaderer, P. B. (2001) "Risk Factors for Asthma Prevalence Among Urban and Nonurban African American Children", *Journal of Asthma*, 38:555-564.
8. Manning, J. P. (1993), *The Role of Histamine, Leukotriene D4 and Inhibitory Prostaglandins in Exercise Bronchoconstriction and Refractoriness in Asthmatic Subjects*, Ph. D. Thesis, Hamilton: McMaster University.
9. Michael T. N., and Peter, J. B. (1998), *Conquering Asthma : An Illustrated*

Guide to Understanding and Care for Adults, Hamilton: B.C. Decker, Inc.

10. Powers, A. D., and Xie, Y. (2000), *Statistical Methods for Categorical Data Analysis*, London: Academic. pp. 88-105.

11. Rees, J., and Kanabar, D. (2000), *ABC of Asthma*, London: BMJ Books.

12 Stokes, M. E., Charles, S. D., and Gray, G. K. (2000), *Categorical Data Analysis Using the SAS System*, North Carolina: SAS Institute, Inc. pp.29-33, 560-568.

13. Tan, A. R., and Spector, L S. (2002), "Exercise-induced Asthma: Diagnosis and Management", *Annals of Allergy, Asthma, and Immunology*, 89:226-238.

APPENDIX A

Details of variable meaning and code for logistic regression analysis

(From the documentation for the NPHS provided by Statistics Canada)

ORIGINAL VARIABLE	FORMATTED VALUE; DATA CLEAR; NEW NAME

Variable Name	PRC6_CUR	**Length**	2	**Position**	7 - 8
Question Name					
Concept	Province of residence				
Question					
Universe	All respondents				
Note	Province of residence at the time of data collection in 1996-1997.				

Content	Code	Sample	Population
NEWFOUNDLAND	10	963	549,322
PRINCE EDWARD ISLAND	11	918	132,322
NOVA SCOTIA	12	986	895,914
NEW BRUNSWICK	13	1,032	728,118
QUÉBEC	24	2,788	7,047,528
ONTARIO	35	39,394	10,839,724
MANITOBA	46	14,828	1,085,635
SASKATCHEWAN	47	1,047	948,511
ALBERTA	48	18,305	2,728,383
BRITISH COLUMBIA	59	1,543	3,686,279
Total		81,804	28,641,736

Formatted value:

$$\begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

PEI, NS, NB, QUE, ONT, MA, SASK, ALB, BC

Variable Name	GE36GURB	**Length**	1	**Position**	9
Question Name					
Concept	Derived rural and urban area - grouped				
Question					
Universe	All respondents				
Note	For sub-provincial analysis, in Ontario, Manitoba and Alberta, use GE36GHRO and in the CMA's of Vancouver and Montréal, use GE36GCMA. See documentation on derived variables.				

Content	Code	Sample	Population
RURAL	1	2,659	2,888,000
URBAN	2	4,903	6,105,200
N/A - ON., MB., AB., VAN. & MTL. CMA	6	74,242	19,648,535
Total		81,804	28,641,735

Formatted value:

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

GE36GU1~2

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME

Variable Name GE36GCMA **Length** 3 **Position** 10 - 12

Question Name

Concept Derived 1991 Census Metropolitan Area - grouped

Question

Universe Respondents living in Montreal/Vancouver

Note See documentation on derived variables.

Content	Code	Sample	Population
MONTRÉAL	462	1,035	3,180,258
VANCOUVER	933	680	1,814,536
NOT APPLICABLE	996	80,089	23,646,941
Total		81,804	28,641,735

Formatted value / data clear / new name: /

Variable Name DHC6_OWN **Length** 1 **Position** 21

Question Name HHLD_Q5

Concept Dwelling owned by household member

Question Is this dwelling owned by a member of this household (even if being paid for)?

Universe All respondents

Note

Content	Code	Sample	Population
YES	1	59,492	20,808,158
NO	2	21,632	7,639,876
DONT KNOW	7	75	25,237
REFUSAL	8	584	158,480
NOT STATED	9	21	9,985
Total		81,804	28,641,736

Formatted value / data clear / new name:

(0, 1)

DHC6_OWN<=2

DHC61~2

Variable Name DHC6GBED **Length** 2 **Position** 22 - 23

Question Name

Concept Derived number of bedrooms in dwelling - grouped

Question

Universe All respondents

Note Based on DHC6_BED.

Content	Code	Sample	Population
NO BEDROOMS	0	326	116,393
1 BEDROOM	1	7,217	1,919,568
2 BEDROOMS	2	17,176	5,417,901
3 BEDROOMS	3	36,206	12,928,301
4 BEDROOMS	4	14,961	6,003,835
5 BEDROOMS OR MORE	5	4,335	1,866,890
NOT STATED	99	1,583	388,847
Total		81,804	28,641,735

Formatted value / data clear / new name:

(0, 1, 2, ⋮, 5)

DHC6GBED<=5

DHC6GB1

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME

Variable Name GE36GHLR **Length** 4 **Position** 13 - 16

Question Name

Concept Derived health areas - 26 groups - grouped

Question

Universe Respondents in Ontario, Manitoba and Alberta

Note See documentation on derived variables.

Content	Code	Sample	Population
NORTH AND SOUTH EASTMAN	461	2,953	86,530
BURNTWOOD, NORMAN, PARKLAND	462	3,059	78,881
CENTRAL, INTERLAKE	463	2,803	153,601
SOUTH WESTMAN, BRANDON, MARQUETTE	464	4,050	111,449
WINNIPEG	465	1,963	655,174
NORTHERN ALBERTA	481	4,488	418,276
SOUTHERN ALBERTA	482	3,070	335,724
CENTRAL ALBERTA	483	3,927	427,650
CALGARY	484	3,694	842,824
EDMONTON	485	3,126	703,909
OTT-CAR.,PRES-RUS.,STOR.,DUN.,GLEN.,RENF	3511	3,824	1,010,980
LAN.,LE.,GREN.,HAST.,P.E.,FRON.,LEN.,ADD	3512	1,978	486,629
N'UMBERLAND, VICT., HALIB.,PETERB., DUR.	3521	3,131	759,832
PEEL, HALTON	3522	3,221	1,243,008
METRO TORONTO	3523	4,085	2,238,057
YORK, SIMCOE	3524	3,224	947,038
NIAGARA	3531	1,506	406,163
HAMILTON-WENTWORTH	3532	1,631	484,221
BRANT, HALDIMAN, NORFOLK	3533	1,534	205,422
WELLINGTON, DUFFERIN, WATERLOO	3534	3,101	634,539
ESSEX, LAMBTON, KENT	3541	3,032	590,194
ELGIN, MIDDLESEX, OXFORD	3542	1,668	609,183
BRUCE, GREY, PERTH, HURON	3543	1,573	297,485
ALGOMA, COCHRANE, MANITOULIN, SUDBURY	3551	2,952	443,511
TIMISKAM., MUSKOKA, PARRY SND., NIPISS.	3552	1,465	233,708
THUNDER BAY, KENORA, RAINY RIVER	3561	1,449	249,755
NOT APPLICABLE	9996	9,277	13,987,994
Total		81,804	28,641,736

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME

Variable Name GE36GHRO **Length** 4 **Position** 17 - 20

Question Name

Concept Derived health areas - 33 groups - grouped

Question

Universe Respondents in Ontario, Manitoba and Alberta

Note These health areas, created in consultation with the provinces for public use purposes, do not correspor to official provincial health regions. See documentation on derived variables.

Content	Code	Sample	Population
NORTH AND SOUTH EASTMAN	461	2,953	86,530
BURNTWOOD, NORMAN, PARKLAND	462	3,059	78,881
CENTRAL, INTERLAKE	463	2,803	153,601
SOUTH WESTMAN, BRANDON, MARQUETTE	464	4,050	111,449
WINNIPEG	465	1,963	655,174
NORTHERN ALBERTA	481	4,488	418,276
SOUTHERN ALBERTA	482	3,070	335,724
CENTRAL ALBERTA	483	3,927	427,650
CALGARY	484	3,694	842,824
EDMONTON	485	3,126	703,909
OTTAWA CARLETON	3511	2,650	738,276
PRES-RUS.,STOR.,DUN.,GLEN.,RENF.	3512	1,174	272,703
LAN, LE,GREN,HAST,P.E, FRONT, LENX, ADD	3513	1,978	486,629
N'UMBERLAND, VICT., HALIB., PETERB.	3521	1,513	277,441
DURHAM	3522	1,618	482,391
PEEL	3523	1,728	882,317
METRO TORONTO	3524	4,085	2,238,057
YORK	3525	1,582	607,098
SIMCOE	3526	1,642	339,940
HALTON	3527	1,493	360,691
NIAGARA	3531	1,506	406,163
HAMILTON-WENTWORTH	3532	1,631	484,221
BRANT, HALDIMAN, NORFOLK	3533	1,534	205,422
WELLINGTON, DUFFERIN	3534	1,546	250,593
WATERLOO	3536	1,555	383,946
ESSEX	3541	1,558	339,105
LAMBTON, KENT	3542	1,474	251,089
ELGIN, MIDDLESEX, OXFORD	3543	1,688	609,183
BRUCE, GREY, PERTH, HURON	3544	1,573	297,485
ALGOMA, COCHRANE	3551	1,473	237,354
MANITOULIN, SUDBURY	3552	1,479	206,157

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME

Variable Name DHC6GAGE **Length** 2 **Position** 37 - 38

Question Name

Concept Age - grouped

Question

Universe All respondents

Note Based on DHC6_AGE.

Content	Code	Sample	Population
0 TO 3 YEARS	1	2,452	807,493
4 TO 5 YEARS	2	1,355	860,745
6 TO 9 YEARS	3	2,867	1,670,452
10 TO 11 YEARS	4	1,728	708,160
12 TO 14 YEARS	5	2,518	1,150,763
15 TO 19 YEARS	6	4,449	2,110,798
20 TO 24 YEARS	7	5,153	1,872,592
25 TO 29 YEARS	8	6,362	2,008,407
30 TO 34 YEARS	9	7,854	2,463,107
35 TO 39 YEARS	10	8,021	2,739,289
40 TO 44 YEARS	11	6,663	2,498,355
45 TO 49 YEARS	12	5,642	2,044,499
50 TO 54 YEARS	13	5,073	1,726,005
55 TO 59 YEARS	14	4,313	1,402,356
60 TO 64 YEARS	15	3,991	1,162,608
65 TO 69 YEARS	16	3,993	1,133,537
70 TO 74 YEARS	17	3,787	962,282
75 TO 79 YEARS	18	2,719	681,003
80 YEARS OR OLDER	19	2,864	639,287
Total		81,804	28,641,736

Formatted value: (1, 2, ⋮, ⋮, 19)

(TREAT AS CONTINUOUS VARIABLE)

DHC6GAGE

Variable Name DHC6_SEX **Length** 1 **Position** 39

Question Name

Concept Sex

Question

Universe All respondents

Note

Content	Code	Sample	Population
MALE	1	38,521	14,170,683
FEMALE	2	43,283	14,471,052
Total		81,804	28,641,735

Formatted value: (0, 1)

DHC6_S1~2

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME
Variable Name DHC6GMAR **Length** 1 **Position** 40 **Question Name** **Concept** Marital status - grouped **Question** **Universe** All respondents **Note** Based on DHC6_MAR. **Content** / **Code** / **Sample** / **Population** MARRIED / COMMON-LAW / PARTNER / 1 / 39,937 / 14,051,736 SINGLE / 2 / 28,616 / 11,561,153 WIDOWED / SEPARATED / DIVORCED / 3 / 13,090 / 2,998,280 NOT STATED / 9 / 161 / 30,566 **Total** / 81,804 / 28,641,735	$\begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$ DHC6GMAR<=3 DHC6GM1~2
Variable Name DHC6GHSZ **Length** 1 **Position** 41 **Question Name** **Concept** Derived household size - grouped **Question** **Universe** All respondents **Note** Based on DHC6DHSZ. **Content** / **Code** / **Sample** / **Population** 1 PERSON / 1 / 16,028 / 3,171,363 2 PERSONS / 2 / 23,727 / 7,201,792 3 PERSONS / 3 / 14,358 / 5,391,788 4 PERSONS / 4 / 16,942 / 7,473,430 5 OR MORE PERSONS / 5 / 10,749 / 5,403,362 **Total** / 81,804 / 28,641,735	$\begin{pmatrix} 1 \\ 2 \\ \vdots \\ \vdots \\ 5 \end{pmatrix}$ DHC6GH
Variable Name DHC6GLE5 **Length** 1 **Position** 42 **Question Name** **Concept** Derived persons <= 5 years old in household - grouped **Question** **Universe** All respondents **Note** Based on DHC6_AGE. **Content** / **Code** / **Sample** / **Population** YES / 1 / 16,460 / 6,284,053 NO / 2 / 65,344 / 22,357,682 **Total** / 81,804 / 28,641,735	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ DHC6GLE1

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME

Variable Name	DHC6G611	**Length**	1	**Position**	43
Question Name					
Concept	Derived persons 6 to 11 years old in hhld - grouped				
Question					
Universe	All respondents				
Note	Based on DHC6_AGE.				

Content	Code	Sample	Population
YES	1	18,209	7,576,504
NO	2	63,595	21,065,231
Total		81,804	28,641,735

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

DHC6G61

Variable Name	DHC6GECF	**Length**	2	**Position**	44 - 45
Question Name					
Concept	Derived household type - grouped				
Question					
Universe	All respondents				
Note	Based on the relationship matrix. See documentation on derived variables.				

Content	Code	Sample	Population
COUPLE WITH CHILDREN < 25	1	31,062	13,239,101
COUPLE W/WO CHILDREN >= 25, W/WO OTHERS	2	2,772	1,247,518
SINGLE	3	16,028	3,171,363
SINGLE WITH OTHERS	4	2,417	787,252
COUPLE WITH CHILDREN < 25, OTHERS	5	3,198	1,694,119
COUPLE ALONE	6	17,967	5,384,483
SINGLE PARENT, CHILDREN < 25	7	5,535	2,068,694
OTHER SINGLE PARENT HOUSEHOLD	8	2,212	840,530
OTHER	9	597	204,347
NOT STATED	99	16	4,330
Total		81,804	28,641,737

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

DHC6GECF<=9

DHC6GE1~6, 8~9

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME

Variable Name DHC6DLVG **Length** 2 **Position** 46 - 47

Question Name

Concept Derived living arrangements of the selected respondent

Question

Universe All respondents

Note Based on the relationship matrix. See documentation on derived variables.

Content	Code	Sample	Population
UNATTACHED INDIVIDUAL LIVING ALONE	1	16,028	3,171,363
UNATTACHED INDIVIDUAL LIVING WITH OTHERS	2	2,211	704,020
LIVING WITH SPOUSE / PARTNER	3	17,966	5,384,406
PARENT LIVING WT SPOUSE/PARTNER,CHILDREN	4	19,818	7,632,923
SINGLE PARENT LIVING WITH CHILDREN	5	3,566	1,043,680
CHILD LIVING WITH SINGLE PARENT	6	1,525	604,825
CHILD LIVING WT SINGLE PARENT, SIBLINGS	7	1,707	891,491
CHILD LIVING WITH TWO PARENTS	8	3,157	1,085,852
CHILD LIVING WITH TWO PARENTS, SIBLINGS	9	10,174	5,434,898
OTHER	10	5,640	2,685,570
NOT STATED	99	12	2,707
Total		81,804	28,641,735

Formatted value / new name:

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

DHC6DLVG<= 10

DHC6DL1~6, 8~9

Variable Name CCC6_1C **Length** 1 **Position** 165

Question Name CHR-Q1

Concept Has asthma

Question Do you have asthma diagnosed by a health professional?

Universe All respondents

Note

Content	Code	Sample	Population
YES	1	6,242	2,236,139
NO	2	75,512	26,394,976
DON'T KNOW	7	38	7,859
REFUSAL	8	5	523
NOT STATED	9	7	2,239
Total		81,804	28,641,736

Respond Variable

CCC6_1C<=2

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME

Variable Name	CCC6_1H	**Length**	1	**Position**	184
Question Name	CHR-Q1				
Concept	Has chronic bronchitis or emphysema				
Question	Do you have chronic bronchitis or emphysema diagnosed by a health professional?				
Universe	All respondents				
Note					

Content	Code	Sample	Population
YES	1	2,603	835,012
NO	2	79,147	27,795,358
DON'T KNOW	7	31	6,912
REFUSAL	8	5	1,235
NOT STATED	9	18	3,219
Total		81,804	28,641,736

Variable Name	SDC6GCB	**Length**	2	**Position**	207 - 208
Question Name					
Concept	Country of birth - grouped				
Question					
Universe	All respondents				
Note	Based on SDC6_1.				

Content	Code	Sample	Population
CANADA	1	69,628	23,771,682
U.S., EUROPE, AUSTRALIA	2	8,353	2,663,564
ASIA	3	2,203	1,418,177
OTHER	4	1,479	737,741
NOT STATED	99	141	50,571
Total		81,804	28,641,735

Formatted value:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

SDC6GCB<=4

SDCGG1~3

Variable Name	SDC6GRAC	**Length**	2	**Position**	213 - 214
Question Name					
Concept	Derived race or colour - grouped				
Question					
Universe	All respondents				
Note	Based on SDC6_7A to SDC6_7L.				

Content	Code	Sample	Population
WHITE	1	75,423	25,461,656
OTHER	2	6,022	3,079,856
NOT STATED	99	359	100,224
Total		81,804	28,641,736

Formatted value:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

SDC6GRAC<=2

SDC6GR1

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME

Variable Name HWC6G3KG **Length** 4 **Position** 290 - 293

Question Name

Concept Weight in kilograms - grouped

Question

Universe All respondents

Note Based on HWC6_3KG. Males >= 137 kgs were grouped into 137 kgs. Females >= 113 were grouped into 113 kgs. Males aged 20 & over who were < 50 kgs were grouped into 50 kgs. Females aged 20 & over who were < 40 kgs were grouped into 40 kgs.

Content	Code	Sample	Population
KILOGRAMS	3 - 137	78,343	27,690,651
NOT STATED	999	3,461	951,084
Total		81,804	28,641,735

Formatted value:

$$\begin{pmatrix} 3 \\ 4 \\ \vdots \\ 137 \end{pmatrix}$$

Data clear: HWC6G3KG <=137

New name: HWC6G3KG

Variable Name LFC6GI13 **Length** 2 **Position** 227 - 228

Question Name

Concept Industry Codes for main job - 13 groups - grouped

Question

Universe Respondents who answered LFC6_2=1

Note Based on LFC6CSIC.

Content	Code	Sample	Population
AGRICULTURAL	1	2,124	404,842
OTHER PRIMARY	2	1,456	362,431
MANUFACTURING / NON-DURABLE	3	2,580	1,097,129
MANUFACTURING / DURABLE	4	2,815	991,389
CONSTRUCTION	5	2,494	831,437
TRANSPORTATION	6	3,310	1,133,835
WHOLESALE TRADE	7	1,652	616,504
RETAIL TRADE	8	5,125	1,890,903
FINANCE	9	2,023	775,239
COMMUNITY SERVICES	10	8,407	2,732,158
PERSONAL SERVICES	11	3,912	1,524,934
BUSINESS SERVICES	12	3,975	1,579,035
PUBLIC ADMINISTRATION	13	2,546	924,500
NOT APPLICABLE	96	36,292	12,927,598
NOT STATED	99	3,093	849,800
Total		81,804	28,641,734

Formatted value:

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

Data clear: LFC6GI13<=96

New name: LFC1~13

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME

Variable Name LFC6GO21 **Length** 2 **Position** 225 - 226

Question Name

Concept Occupation Codes for main job - 21 groups - grouped

Question

Universe Respondents who answered LFC6_2=1

Note Based on LFC6CSOC. See documentation on derived variables.

Content	Code	Sample	Population
MANAGERIAL, ADMINISTRATIVE	1	5,825	2,062,784
NATURAL SCIENCE	2	1,780	601,459
SOCIAL SCIENCE	3	994	358,871
RELIGION	4	121	26,294
TEACHING	5	2,232	740,580
MEDICINE	6	2,338	739,736
ARTISTIC	7	868	319,723
CLERICAL	8	5,694	2,243,093
SALES	9	3,796	1,365,650
SERVICES	10	5,733	2,171,099
FARMING	11	2,111	418,617
FISHING	12	96	53,532
FORESTRY	13	188	85,416
MINING	14	408	53,215
PROCESSING	15	1,011	383,650
MACHINING	16	696	191,581
FABRICATING	17	3,004	1,149,963
CONSTRUCTION	18	2,422	800,219
TRANSPORTATION	19	1,770	564,030
MATERIALS HANDLING	20	888	393,366
OTHER CRAFTS	21	444	141,460
NOT APPLICABLE	96	36,292	12,927,598
NOT STATED	99	3,093	849,800
Total		81,804	28,641,735

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME

Variable Name HWC6GBMI **Length** 4.1 **Position** 296 - 299

Question Name

Concept Derived Body Mass Index (1 decimal place) - grouped

Question

Universe Respondents aged 20 to 64 years old excluding pregnant women

Note Based on HWC6GHT, HWC6G3KG and HWC6_1. See documentation on derived variables.

Content	Code	Sample	Population
BMI SCORE	13.2 - 57.9	50,347	17,164,709
NOT APPLICABLE	99.6	29,474	10,939,358
NOT STATED	99.9	1,983	537,668
Total		81,804	28,641,736

$$\begin{pmatrix} 13.2 \\ \vdots \\ 57.9 \\ 0 \end{pmatrix}$$

HWC6GBMI<=99.6

HWC6GB1

Variable Name HWC6GSW **Length** 1 **Position** 300

Question Name

Concept Derived standard weight - grouped

Question

Universe Respondents aged 20 to 64 years old excluding pregnant women

Note Based on HWC6GBMI. See documentation on derived variables.

Content	Code	Sample	Population
INSUFFICIENT WEIGHT	1	3,805	1,422,385
ACCEPTABLE WEIGHT	2	21,014	7,474,777
SOME EXCESS WEIGHT	3	9,779	3,287,154
OVERWEIGHT	4	15,749	4,980,393
NOT APPLICABLE	6	29,474	10,939,358
NOT STATED	9	1,983	537,668
Total		81,804	28,641,735

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

HWC6GSW<=6

HWC6G1~4

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME
Variable Name HWC6GHT **Length** 2 **Position** 288 - 289 **Question Name** **Concept** Height - adults and children - grouped **Question** **Universe** All respondents **Note** Based on HWC6_HT. Males > 6'2" were grouped into 6'2". Females > 5'11" were grouped into 5'11". Males aged 20 + who were < 5' were grouped into 5'. Females aged 20 + who were < 4'7" were groupe into 4'7".	(2 : : : : 64) HWC6GHT<=64

Content	Code	Sample	Population
1'0" (12 INCHES) (29.2 TO 31.7 CM)	2	1	113
1'2" (14 INCHES) (34.3 TO 36.7 CM)	4	1	160
1'3" (15 INCHES) (36.8 TO 39.3 CM)	5	2	613
1'4" (16 INCHES) (39.4 TO 41.8 CM)	6	3	490
1'5" (17 INCHES) (41.9 TO 44.4 CM)	7	2	234
1'6" (18 INCHES) (44.5 TO 46.9 CM)	8	10	4,346
1'7" (19 INCHES) (47.0 TO 49.4 CM)	9	4	306
1'8" (20 INCHES) (49.5 TO 52.0 CM)	10	16	1,588
1'9" (21 INCHES) (52.1 TO 54.5 CM)	11	26	2,781
1'10" (22 INCHES) (54.6 TO 57.1 CM)	12	39	3,622
1'11" (23 INCHES) (57.2 TO 59.6 CM)	13	54	4,351
2'0" (24 INCHES) (59.7 TO 62.1 CM)	14	133	36,978
2'1" (25 INCHES) (62.2 TO 64.7 CM)	15	52	7,319
2'2" (26 INCHES) (64.8 TO 67.2 CM)	16	84	10,392
2'3" (27 INCHES) (67.3 TO 69.8 CM)	17	54	11,098
2'4" (28 INCHES) (69.9 TO 72.3 CM)	18	81	14,271
2'5" (29 INCHES) (72.4 TO 74.8 CM)	19	89	15,413
2'6" (30 INCHES) (74.9 TO 77.4 CM)	20	208	67,043
2'7" (31 INCHES) (77.5 TO 79.9 CM)	21	61	19,928
2'8" (32 INCHES) (80.0 TO 82.5 CM)	22	101	19,085
2'9" (33 INCHES) (82.6 TO 85.0 CM)	23	94	25,673
2'10" (34 INCHES) (85.1 TO 87.5 CM)	24	61	12,939
2'11" (35 INCHES) (87.6 TO 90.1 CM)	25	83	43,312
3'0" (36 INCHES) (90.2 TO 92.6 CM)	26	643	214,957
3'1" (37 INCHES) (92.7 TO 95.2 CM)	27	112	65,670
3'2" (38 INCHES) (95.3 TO 97.7 CM)	28	154	100,600
3'3" (39 INCHES) (97.8 TO 100.2 CM)	29	110	80,384
3'4" (40 INCHES) (100.3 TO 102.8 CM)	30	199	129,059
3'5" (41 INCHES) (102.9 TO 105.3 CM)	31	171	103,909
3'6" (42 INCHES) (105.4 TO 107.9 CM)	32	391	255,468
3'7" (43 INCHES) (108.0 TO 110.4 CM)	33	115	96,566

ORIGINAL VARIABLE				FORMATTED VALUE DATA CLEAR NEW NAME
3'8" (44 INCHES) (110.5 TO 112.9 CM)	34	136	98,490	
3'9" (45 INCHES) (113.0 TO 115.5 CM)	35	133	89,220	
3'10" (46 INCHES) (115.6 TO 118.0 CM)	36	97	68,852	
3'11" (47 INCHES) (118.1 TO 120.6 CM)	37	116	98,204	
4'0" (48 INCHES) (120.7 TO 123.1 CM)	38	1,026	576,944	
4'1" (49 INCHES) (123.2 TO 125.6 CM)	39	152	122,648	
4'2" (50 INCHES) (125.7 TO 128.2 CM)	40	251	171,253	
4'3" (51 INCHES) (128.3 TO 130.7 CM)	41	162	124,988	
4'4" (52 INCHES) (130.8 TO 133.3 CM)	42	155	138,897	
4'5" (53 INCHES) (133.4 TO 135.8 CM)	43	201	120,695	
4'6" (54 INCHES) (135.9 TO 138.3 CM)	44	339	191,507	
4'7" (55 INCHES) (138.4 TO 140.9 CM)	45	218	103,259	
4'8" (56 INCHES) (141.0 TO 143.4 CM)	46	298	136,943	
4'9" (57 INCHES) (143.5 TO 146.0 CM)	47	260	115,720	
4'10" (58 INCHES) (146.1 TO 148.5 CM)	48	503	195,802	
4'11" (59 INCHES) (148.6 TO 151.0 CM)	49	825	336,618	
5'0" (60 INCHES) (151.1 TO 153.6 CM)	50	2,909	1,087,011	
5'1" (61 INCHES) (153.7 TO 156.1 CM)	51	2,748	954,010	
5'2" (62 INCHES) (156.2 TO 158.7 CM)	52	5,674	1,843,299	
5'3" (63 INCHES) (158.8 TO 161.2 CM)	53	5,249	1,847,610	
5'4" (64 INCHES) (161.3 TO 163.7 CM)	54	6,712	2,118,939	
5'5" (65 INCHES) (163.8 TO 166.3 CM)	55	5,566	1,952,506	
5'6" (66 INCHES) (166.4 TO 168.8 CM)	56	6,593	2,209,327	
5'7" (67 INCHES) (168.9 TO 171.4 CM)	57	6,193	2,090,665	
5'8" (68 INCHES) (171.5 TO 173.9 CM)	58	6,499	2,107,348	
5'9" (69 INCHES) (174.0 TO 176.4 CM)	59	4,576	1,574,484	
5'10" (70 INCHES) (176.5 TO 179.0 CM)	60	5,536	1,958,810	
5'11" (71 INCHES) (179.1 TO 181.5 CM)	61	4,823	1,533,474	
6'0" (72 INCHES) (181.6 TO 184.1 CM)	62	4,194	1,294,319	
6'1" (73 INCHES) (184.2 TO 186.6 CM)	63	1,837	548,644	
6'2" (74 INCHES) (186.7 TO 189.1 CM)	64	2,864	918,287	
NOT STATED	99	1,805	564,298	
Total		81,804	28,641,736	

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME

Variable Name PAC6DLEI **Length** 1 **Position** 796

Question Name

Concept Derived participant in leisure physical activity

Question

Universe Respondents aged 12 and over

Note Based on PAC6_1A to PAC6_1X.

Content	Code	Sample	Population
YES	1	63,304	21,280,090
NO	2	8,073	2,555,779
NOT APPLICABLE	6	8,402	4,046,849
NOT STATED	9	2,025	759,018
Total		81,804	28,641,736

Formatted value: $\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$

Data clear: PAC6DLEI<=6

New name: PAC6DL1

Variable Name PAC6DFM **Length** 3 **Position** 797 - 799

Question Name

Concept Derived monthly freq. of physical activity lasting > 15 min.

Question

Universe Respondents aged 12 and over

Note Based on PAC6_2A to PAC6_2X. See documentation on derived variables.

Content	Code	Sample	Population
MONTHLY FREQUENCY	0 - 255	71,377	23,835,869
NOT APPLICABLE	996	8,402	4,046,849
NOT STATED	999	2,025	759,018
Total		81,804	28,641,735

Formatted value: $\begin{pmatrix} 0.01 \\ 1 \\ \vdots \\ 255 \\ 0 \end{pmatrix}$

Data clear: PAC6DFM<=996

New name: PAC6DFM1

Variable Name PAC6DEE **Length** 4.1 **Position** 792 - 795

Question Name

Concept Derived energy expenditure (1 decimal point)

Question

Universe Respondents aged 12 and over

Note Based on PAC6_1A to PAC6_1X and PAC6_2. See documentation on derived variables.

Content	Code	Sample	Population
AMOUNT OF ENERGY EXPENDITURE	0.0 - 34.4	71,377	23,835,869
NOT APPLICABLE	99.6	8,402	4,046,849
NOT STATED	99.9	2,025	759,018
Total		81,804	28,641,735

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME

Variable Name PAC6DFR **Length** 1 **Position** 800

Question Name

Concept Derived frequency of all physical activity

Question

Universe Respondents aged 12 and over

Note Based on PAC6DFM. See documentation on derived variables.

Content	Code	Sample	Population
REGULAR	1	42,545	14,156,094
OCCASIONAL	2	12,749	4,536,763
INFREQUENT	3	16,083	5,143,012
NOT APPLICABLE	6	8,402	4,046,849
NOT STATED	9	2,025	759,018
	Total	81,804	28,641,736

Formatted value / data clear / new name: /

Variable Name PAC6DFD **Length** 1 **Position** 801

Question Name

Concept Derived participation / daily phys. activities > 15 min.

Question

Universe Respondents aged 12 and over

Note Based on PAC6DFM. See documentation on derived variables.

Content	Code	Sample	Population
DAILY	1	22,601	7,305,908
NOT DAILY	2	48,776	16,529,961
NOT APPLICABLE	6	8,402	4,046,849
NOT STATED	9	2,025	759,018
	Total	81,804	28,641,736

Formatted value / data clear / new name: /

Variable Name PAC6DPAI **Length** 1 **Position** 802

Question Name

Concept Derived physical activity index

Question

Universe Respondents aged 12 and over

Note Based on PAC6DEE. See documentation on derived variables.

Content	Code	Sample	Population
ACTIVE	1	14,377	4,911,009
MODERATE	2	16,346	5,382,985
INACTIVE	3	40,654	13,541,875
NOT APPLICABLE	6	8,402	4,046,849
NOT STATED	9	2,025	759,018
	Total	81,804	28,641,736

Formatted value / data clear / new name:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

PAC6DPAI<=6

ORIGINAL VARIABLE	FORMATTED VALUE DATA CLEAR NEW NAME

Variable Name SMC6_3 **Length** 3 **Position** 927 - 929

Question Name SMK-Q3

Concept Age started smoking daily - daily smoker

Question At what age did you begin to smoke cigarettes daily?

Universe Respondents who answered SMC6_2=1

Note

Content	Code	Sample	Population
YEARS	5 - 85	17,248	5,718,347
NOT APPLICABLE	996	64,077	22,778,375
DON'T KNOW	997	194	56,590
REFUSAL	998	33	19,581
NOT STATED	999	252	68,842
Total		81,804	28,641,735

Formatted value: $\begin{pmatrix} 5 \\ \vdots \\ 85 \\ 100 \end{pmatrix}$

SMC6_3<=996

SMC6_31

Variable Name SMC6_4 **Length** 3 **Position** 933 - 935

Question Name SMK-Q4

Concept Number of cigarettes smoked each day - daily smoker

Question How many cigarettes do you smoke each day now?

Universe Respondents who answered SMC6_2=1

Note

Content	Code	Sample	Population
NUMBER OF CIGARETTES	1 - 99	17,354	5,767,704
NOT APPLICABLE	996	64,077	22,778,375
DON'T KNOW	997	88	21,936
REFUSAL	998	33	4,878
NOT STATED	999	252	68,842
Total		81,804	28,641,735

Formatted value: $\begin{pmatrix} 1 \\ \vdots \\ 99 \\ 0 \end{pmatrix}$

SMC6_4<=996

SMC6_41

Variable Name WT66 **Length** 8.2 **Position** 1228 - 1235

Question Name

Concept Sampling weight for selected respondent

Question

Universe All respondents

Note See documentation on weighting.

Weight used in all analysis

APPENDIX B

Results of contingency table analysis.

Variable Name & Meaning	χ^2 test	d.f.	Statistic value	P value
PRC6_CUR (Province of residence)	Pearson Chi-Square LLR Chi-Square	9 9	22.3672 23.4992	0.0078 0.0052
GE36GURB (Derived rural and urban area – grouped)	Pearson Chi-Square LLR Chi-Square	2 2	21.2294 22.2515	<.0001 <.0001
GE36GCMA (Derived 1991 Census Metropolitan Area – grouped)	Pearson Chi-Square LLR Chi-Square	2 2	4.1918 4.1365	0.1230 0.1264
GE36GHLR (Derived health areas – 26 groups – grouped)	Pearson Chi-Square LLR Chi-Square	26 26	103.1537 99.6342	<.0001 <.0001
GE36GHRO (Derived health areas – 33 groups – grouped)	Pearson Chi-Square LLR Chi-Square	33 33	122.2269 116.5697	<.0001 <.0001
DHC6GAGE (Age – grouped)	Pearson Chi-Square LLR Chi-Square	18 18	1012.3583 943.1849	<.0001 <.0001
DHC6_SEX (Gender)	Pearson Chi-Square LLR Chi-Square	1 1	50.2127 50.2966	<.0001 <.0001
DHC6GHSZ (Derived household size – grouped)	Pearson Chi-Square LLR Chi-Square	4 4	5.9336 5.8921	0.2042 0.2074
DHC6GLE5 (Derived persons <= 5 years old in household – grouped)	Pearson Chi-Square LLR Chi-Square	1 1	0.3673 0.3662	0.5445 0.5451
DHC6G611 (Derived persons 6 to 11 years old in hhld – grouped)	Pearson Chi-Square LLR Chi-Square	1 1	36.6438 35.8356	<.0001 <.0001
HWC6G3KG (Height – adults and children – grouped)	Pearson Chi-Square LLR Chi-Square	13 13	394.9805 380.9162	<.0001 <.0001
DHC6DLVG (Derived living arrangements of the selected respondent)	Pearson Chi-Square LLR Chi-Square	9 9	610.7598 608.6087	<.0001 <.0001
HWC6GSW (Weight in kilograms - grouped)	Pearson Chi-Square LLR Chi-Square	4 4	398.7164 393.4118	<.0001 <.0001
HWC6GBMI (Derived Body Mass Index (1 decimal place) - grouped)	Pearson Chi-Square LLR Chi-Square	5 5	447.3341 439.5541	<.0001 <.0001

LFC6GO21 (Occupation Codes for main job - 21 groups – grouped)	Pearson Chi-Square LLR Chi-Square	21 21	395.7104 425.7088	<.0001 <.0001
LFC6GI13 (Industry Codes for main job - 13 groups – grouped)	Pearson Chi-Square LLR Chi-Square	13 13	312.7330 335.4328	<.0001 <.0001
PAC6DEE (Derived energy expenditure (1 decimal point))	Pearson Chi-Square LLR Chi-Square	4 4	222.8556 203.3757	<.0001 <.0001
PAC6DLEI (Derived participant in leisure physical activity)	Pearson Chi-Square LLR Chi-Square	2 2	228.8690 209.4993	<.0001 <.0001
PAC6DFM (Derived monthly freq. of physical activity lasting > 15 min.)	Pearson Chi-Square LLR Chi-Square	6 6	300.9466 286.6957	<.0001 <.0001
PAC6DFR (Derived frequency of all physical activity)	Pearson Chi-Square LLR Chi-Square	3 3	254.7385 237.6286	<.0001 <.0001
PAC6DFD (Derived participation / daily phys. activities > 15 min.)	Pearson Chi-Square LLR Chi-Square	2 2	292.1264 275.1194	<.0001 <.0001
PAC6DPAI (Derived physical activity index)	Pearson Chi-Square LLR Chi-Square	3 3	306.5320 290.5368	<.0001 <.0001
SDC6GCB (Country of birth - grouped)	Pearson Chi-Square LLR Chi-Square	3 3	387.1970 498.4626	<.0001 <.0001
SDC6GRAC (Derived race or colour – grouped)	Pearson Chi-Square LLR Chi-Square	1 1	38.9072 41.4480	<.0001 <.0001
SMC6_4 (Number of cigarettes smoked each day - daily smoker)	Pearson Chi-Square LLR Chi-Square	54 54	97.8939 94.8793	0.0002 0.0005
CCC6_1H (Has chronic bronchitis or emphysema)	Pearson Chi-Square LLR Chi-Square	1 1	2316.2080 1391.1395	<.0001 <.0001
DHC6GBED (Derived number of bedrooms in dwelling – grouped)	Pearson Chi-Square LLR Chi-Square	5 5	41.8390 40.9440	<.0001 <.0001
DHC6GECF (Derived household type – grouped)	Pearson Chi-Square LLR Chi-Square	8 8	235.6656 231.6044	<.0001 <.0001
HC6GMAR (Marital status – grouped)	Pearson Chi-Square LLR Chi-Square	2 2	688.2057 687.5607	<.0001 <.0001
HWC6GHT (Derived standard weight – grouped)	Pearson Chi-Square LLR Chi-Square	6 6	383.8739 355.0966	<.0001 <.0001

Appendix C

Program used to extract data of NPHS from H356.txt in abbreviated form.

```
proc format;
    VALUE    ACCQ71FM
           1      = '< THAN 1 YR AGO'
           2      = '1 TO <2YEARS AGO'
           3      = '2 - <3 YEARS AGO'
           4      = '3 OR + YEARS AGO'
           5      = 'NEVER'
           6      = 'NOT APPLICABLE'
           7      = 'DON''T KNOW'
           8      = 'REFUSAL'
           9      = 'NOT STATED'
                        ;
    VALUE    ACCQ72FM
           1      = '> THAN ONCE A YR'
           2      = 'ONCE A YEAR'
           3      = 'EVERY 2 YEARS'
           4      = 'EVERY 3 YEARS'
           5      = '<THAN EVERY 3 YR'
           6      = 'NOT APPLICABLE'
           7      = 'DON''T KNOW'
           8      = 'REFUSAL'
           9      = 'NOT STATED'
                        ;
          :
          :
          :
                        ;
    VALUE    YEARFM
        9996      = 'NOT APPLICABLE'
        9997      = 'DON''T KNOW'
        9998      = 'REFUSAL'
        9999      = 'NOT STATED'
                        ;
    VALUE    YESNOFM
           1      = 'YES'
           2      = 'NO'
           6      = 'NOT APPLICABLE'
           7      = 'DON''T KNOW'
           8      = 'REFUSAL'
           9      = 'NOT STATED'
                        ;
run;
data 'd:\public\h962';
INFILE 'E:\SAS\H356.txt' LRECL = 1243;
INPUT
    @   1    AM66_RNO       6.
    @   7    PRC6_CUR       2.
    @   9    GE36GURB       1.
    @  10    GE36GCMA       3.
```

```
    @  13    GE36GHLR       4.
    @  17    GE36GHRO       4.
    @  21    DHC6_OWN       1.
    @  22    DHC6GBED       2.
    @  24    INC6G2         2.
    @  26    INC6DIA5       1.
            :
            :
            :
    @1222    VSP6_7         1.
    @1223    VSP6_8         1.
    @1224    VSP6_9         1.
    @1225    VSP6_10        1.
    @1226    COP6_1         1.
    @1227    COP6_2         1.
    @1228    WT66           8.2
    @1236    WT66_N         8.2
         ;
label
    AM66_RNO  =  'Record number on Health Microdata file'
    PRC6_CUR  =  'Province of residence'
    GE36GURB  =  'Derived rural and urban area - grouped'
    GE36GCMA  =  'Derived 1991 Census Metropolitan Area - grouped'
          :
          :
          :
    VSP6_7    =  'Number of times attacked at school / school bus'
    VSP6_8    =  'Number of times verbally abused outside of school'
    WT66      =  'Sampling weight for selected respondent'
                             ;
 format
    PRC6_CUR    PROVFM.
    GE36GURB    DVGURBFM.
    GE36GCMA    DVGCMAFM.
    GE36GHLR    DVGHLRFM.
        :
        :
        :
    RSS6_9      YESNOFM.
    RSS6_10     RSS_FREQ.
    VSP6_1      OFTENFM.
    VSP6_2      OFTENFM.
    VSP6_4      OFTENFM.
    VSP6_5      VSP_NUM.
    COP6_1      GOOD5FM.
    COP6_2      GOOD5FM.
                       ;
run;
```

APPENDIX D

Programs used for contingency table analysis.

Program 1.

```
data 'e:\temp4\data_a';
Set  'd:\public\h962';
keep wt66;
if ccc6_1c<=2;
run;
proc means data='e:\temp4\data_a' noprint;
var WT66;
output out='e:\temp4\aaa' mean=mwt66;
run;
proc print;
run;
```

Program 2.

```
data 'e:\temp4\data';
Set  'd:\public\h962';
Keep
PRC6_CUR
GE36GURB
GE36GCMA
GE36GHLR
GE36GHRO
DHC6_OWN
DHC6GBED
DHC6GAGE
DHC6_SEX
DHC6GMAR
DHC6GHSZ
DHC6GLE5
DHC6G611
DHC6GECF
DHC6DLVG
CCC6_1C
CCC6_C5
CCC6_C6
CCC6_1H
SDC6GCB
SDC6GRAC
LFC6GO21
LFC6GI13
HWC6GHT
HWC6G3KG
HWC6GBMI
HWC6GSW
HWS6_1
HWS6_5
PAC6DEE
PAC6DLEI
PAC6DFM
PAC6DFR
PAC6DFD
PAC6DPAI
DGC6_1G
DGK6_1
SMC6_3
SMC6_4;
```

```
Run;
data 'e:\temp4\data';
set 'e:\temp4\data';
if CCC6_1C<=2;
Wt66=Wt66/350.21055985;
run;
```

Program 3.

```
proc freq data='e:\temp4\data';
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables(PRC6_CUR
       GE36GURB
       GE36GCMA
       GE36GHLR
       GE36GHRO
       DHC6GAGE
       DHC6_SEX
       DHC6GHSZ
       DHC6GLE5
       DHC6G611)*CCC6_1C/chisq ;
*output out='e:\temp1\tfreq0';
run;

options nonumber nodate pagesize=5000;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 HWC6G3KG;
if HWC6G3KG<=140;
if 0<=HWC6G3KG<=10 then HWC6G3KG=1;
if 10<HWC6G3KG<=20 then HWC6G3KG=2;
if 20<HWC6G3KG<=30 then HWC6G3KG=3;
if 30<HWC6G3KG<=40 then HWC6G3KG=4;
if 40<HWC6G3KG<=50 then HWC6G3KG=5;
if 50<HWC6G3KG<=60 then HWC6G3KG=6;
if 60<HWC6G3KG<=70 then HWC6G3KG=7;
if 70<HWC6G3KG<=80 then HWC6G3KG=8;
if 80<HWC6G3KG<=90 then HWC6G3KG=9;
if 90<HWC6G3KG<=100 then HWC6G3KG=10;
if 100<HWC6G3KG<=110 then HWC6G3KG=11;
if 110<HWC6G3KG<=120 then HWC6G3KG=12;
if 120<HWC6G3KG<=130 then HWC6G3KG=13;
if 130<=HWC6G3KG<=140 then HWC6G3KG=14;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables HWC6G3KG*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66  HWC6GBMI ;
if  HWC6GBMI <=99.6;
if 10<HWC6GBMI<=20 then HWC6GBMI=1;
if 20<HWC6GBMI<=30 then HWC6GBMI=2;
```

```
if 30<HWC6GBMI<=40 then HWC6GBMI=3;
if 40<HWC6GBMI<=50 then HWC6GBMI=4;
if 50<HWC6GBMI<=60 then HWC6GBMI=5;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables HWC6GBMI*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66   HWC6GSW ;
if   HWC6GSW <=6;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables HWC6GSW*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66  HWC6GHT ;
if    HWC6GHT <=64;
if 0<=HWC6GHT<=10 then HWC6GHT=1;
if 10<HWC6GHT<=20 then HWC6GHT=2;
if 20<HWC6GHT<=30 then HWC6GHT=3;
if 30<HWC6GHT<=40 then HWC6GHT=4;
if 40<HWC6GHT<=50 then HWC6GHT=5;
if 50<HWC6GHT<=60 then HWC6GHT=6;
if 60<HWC6GHT<=70 then HWC6GHT=7;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables  HWC6GHT*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66  LFC6GO21 ;
if   LFC6GO21 <=96;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables LFC6GO21*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66  LFC6GI13 ;
if   LFC6GI13 <=96;
run;
proc freq data=freq ;
weight WT66;
```

```
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables LFC6GI13*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 PAC6DEE  ;
if   PAC6DEE  <=99.6;
if 0<=PAC6DEE<=10 then PAC6DEE=1;
if 10<PAC6DEE<=20 then PAC6DEE=2;
if 20<PAC6DEE<=30 then PAC6DEE=3;
if 30<PAC6DEE<=40 then PAC6DEE=4;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables PAC6DEE*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66   PAC6DLEI ;
if    PAC6DLEI <=6;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables  PAC6DLEI*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 PAC6DFM ;
if  PAC6DFM <=996;
if 0<=PAC6DFM<=5 then PAC6DFM=1;
if 5<PAC6DFM<=10 then PAC6DFM=2;
if 10<PAC6DFM<=15 then PAC6DFM=3;
if 15<PAC6DFM<=20 then PAC6DFM=4;
if 20<PAC6DFM<=25 then PAC6DFM=5;
if 25<PAC6DFM<=255 then PAC6DFM=6;
if 255<PAC6DFM then PAC6DFM=7;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables  PAC6DFM*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 PAC6DFR ;
if   PAC6DFR <=6;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables  PAC6DFR*CCC6_1C/Chisq;
```

```
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 PAC6DFD;
if PAC6DFD<=6;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables  PAC6DFD*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 PAC6DPAI;
if PAC6DPAI<=6;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables PAC6DPAI*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 SDC6GCB;
if SDC6GCB<=4;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables SDC6GCB*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 SDC6GRAC;
if SDC6GRAC<=2;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables SDC6GRAC*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 SMC6_3;
if SMC6_3<=996;
if 0<=SMC6_3<=10 then SMC6_3=1;
if 10<SMC6_3<=20 then SMC6_3=2;
if 20<SMC6_3<=30 then SMC6_3=3;
if 30<SMC6_3<=40 then SMC6_3=4;
if 40<SMC6_3<=50 then SMC6_3=5;
if 50<SMC6_3<=60 then SMC6_3=6;
if 60<SMC6_3<=70 then SMC6_3=7;
```

```
if 60<SMC6_3<=70 then SMC6_3=7;
   else SMC6_3=8;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables SMC6_3*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 SMC6_4;
if SMC6_4<=996;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables SMC6_4*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 CCC6_1H;
if CCC6_1H<=2;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables CCC6_1H*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 DHC6GBED;
if 0<=DHC6GBED<=5;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables DHC6GBED*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 DHC6GMAR;
if 0<=DHC6GMAR<=3;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables DHC6GMAR*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 DHC6GECF;
if 0<=DHC6GECF<=9;
```

```
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables DHC6GECF*CCC6_1C/Chisq;
run;

data freq;
set 'e:\temp4\data';
keep CCC6_1C WT66 DHC6DLVG;
if 0<=DHC6DLVG<=10;
run;
proc freq data=freq ;
weight WT66;
title "FREQUENCY TABLE ABOUT VARIABLES RELATED TO ASTHMA";
tables DHC6DLVG*CCC6_1C/Chisq;
run;
```

APPENDIX E

Programs for trend test.

```
data trend;
SET 'e:\temp4\data';
KEEP CCC6_1C WT66 DHC6GAGE ;
RUN;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*DHC6GAGE/trend NOPRINT ;
RUN;
/****************/
data trend;
SET 'e:\temp4\data';
KEEP CCC6_1C WT66 DHC6GBED ;
IF 0<=DHC6GBED<=99;
RUN;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*DHC6GBED/trend NOPRINT ;
RUN;
/***************/
data trend;
SET 'e:\temp4\data';
KEEP CCC6_1C WT66 DHC6GHSZ ;
RUN;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*DHC6GHSZ/trend NOPRINT ;
RUN;
/***************/
data trend;
SET 'e:\temp4\data';
KEEP PAC6DEE CCC6_1C WT66;
IF PAC6DEE<=34.4;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*PAC6DEE/trend noprint;
RUN;
/***************/
data trend;
SET 'e:\temp4\data';
KEEP PAC6DFM CCC6_1C WT66;
IF PAC6DFM<=255;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*PAC6DFM/trend NOPRINT ;
RUN;

/***************/
data trend;
SET 'e:\temp4\data';
```

```
KEEP PAC6DFR CCC6_1C WT66;
IF PAC6DFR<=3;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*PAC6DFR/trend NOPRINT ;
RUN;
/***************/
data trend;
SET 'e:\temp4\data';
KEEP PAC6DPAI CCC6_1C WT66  SMC6_3 ;
IF SMC6_3<=85;
RUN;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*SMC6_3/TREND NOPRINT ;
RUN;
/***************/
data trend;
SET 'e:\temp4\data';
KEEP CCC6_1C WT66  SMC6_4 ;
IF 0<=SMC6_4<=99;
RUN;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*SMC6_4/trend NOPRINT ;
RUN;
/***************/
data trend;
SET 'e:\temp4\data';
KEEP CCC6_1C WT66  SMC6_4 ;
IF 0<=SMC6_4<=15;
RUN;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*SMC6_4/trend NOPRINT ;
RUN;
/***************/
data trend;
SET 'e:\temp4\data';
KEEP CCC6_1C WT66  SMC6_4 ;
IF 15<=SMC6_4<=98;
RUN;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*SMC6_4/trend NOPRINT ;
RUN;
/***************/
data trend;
SET 'e:\temp4\data';
KEEP CCC6_1C WT66  HWC6GHT ;
IF 0<=HWC6GHT<99;
RUN;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*HWC6GHT/trend NOPRINT ;
RUN;
```

```
/***************/
data trend;
SET 'e:\temp4\data';
KEEP CCC6_1C WT66  HWC6GSW ;
IF 0<=HWC6GSW<6;
RUN;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*HWC6GSW/trend NOPRINT ;
RUN;

/***************/
data trend;
SET 'e:\temp4\data';
KEEP CCC6_1C WT66  HWC6GSW ;
IF 0<=HWC6GSW<=3;
RUN;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*HWC6GSW/trend NOPRINT ;
RUN;
/***************/
data trend;
SET 'e:\temp4\data';
KEEP CCC6_1C WT66  HWC6G3KG ;
IF 0<=HWC6G3KG<=137;
RUN;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*HWC6G3KG/trend NOPRINT ;
RUN;
/***************/
data trend;
SET 'e:\temp4\data';
KEEP CCC6_1C WT66  HWC6G3KG ;
IF 70<=HWC6G3KG<=137;
RUN;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*HWC6G3KG/trend NOPRINT ;
RUN;
/***************/
data trend;
SET 'e:\temp4\data';
KEEP CCC6_1C WT66  HWC6GBMI ;
IF 0<=HWC6GBMI<=60;
RUN;
PROC FREQ DATA=TREND;
WEIGHT WT66;
TABLES CCC6_1C*HWC6GBMI/trend NOPRINT ;
RUN;
/***************/
```

APPENDIX F

Program used for odds ratio and relative risk analysis.

```
data data;
set 'e:\temp4\data';
keep CCC6_1C WT66 PRC6_CUR ;
RUN;
/**************
DATA DATA;
SET DATA;
if PRC6_CUR=10 THEN NFLD=0; ELSE NFLD=1;
IF PRC6_CUR=11 THEN PEI=0; ELSE PEI=1;
IF PRC6_CUR=12 THEN NS=0; ELSE NS =1;
IF PRC6_CUR=13 THEN NB=0; ELSE NB=1;
IF PRC6_CUR=24 THEN QUE=0; ELSE QUE=1;
IF PRC6_CUR=35 THEN ONT=0; ELSE ONT=1;
IF PRC6_CUR=46 THEN MB=0; ELSE MB=1;
IF PRC6_CUR=47 THEN SASK=0; ELSE SASK=1;
IF PRC6_CUR=48 THEN AB=0; ELSE AB=1;
IF PRC6_CUR=59 THEN BC=0; ELSE BC=1;
RUN;
PROC FREQ DATA=DATA;
weight WT66;
tables (NFLD PEI NS NB QUE ONT MB SASK AB BC)*CCC6_1C
       /chisq CMH  ALPHA=0.01 noprint;
run;
************************/
DATA TEST;
SET DATA;
IF PRC6_CUR=12 OR PRC6_CUR=10;
IF PRC6_CUR=10 THEN PRC6_CUR=100;
PROC FREQ DATA=TEST;
WEIGHT WT66;
TABLES PRC6_CUR*CCC6_1C/CHISQ MEASURE ALPHA=0.01;
RUN;

DATA TEST;
SET DATA;
IF PRC6_CUR=12 OR PRC6_CUR=35 OR PRC6_CUR=24 THEN HIGH_LOW=0;ELSE
HIGH_LOW=1;
PROC FREQ DATA=TEST;
WEIGHT WT66;
TABLES HIGH_LOW*CCC6_1C/CHISQ MEASURE ALPHA=0.01;
RUN;
************************/

data data;
set 'e:\temp4\data';
keep CCC6_1C WT66 GE36GURB ;
IF GE36GURB<=2;
```

```
RUN;
PROC FREQ DATA=DATA;
weight WT66;
TABLES GE36GURB*CCC6_1C/CHISQ MEASURES ALPHA=0.01;
RUN;

************************/
data data;
set 'e:\temp4\data';
keep CCC6_1C WT66 LFC6GI13 ;
IF LFC6GI13<=13;
RUN;

DATA TEST;
SET DATA;
IF LFC6GI13=1 OR LFC6GI13=11;
IF LFC6GI13=11 THEN LFC6GI13=0;
PROC FREQ DATA=TEST;
WEIGHT WT66;
TABLES LFC6GI13*CCC6_1C/CHISQ MEASURE ALPHA=0.01;
RUN;


DATA TEST;
SET DATA;
IF LFC6GI13=9 OR LFC6GI13=10 OR LFC6GI13=11 THEN HIGH_LOW=0;ELSE
HIGH_LOW=1;
PROC FREQ DATA=TEST;
WEIGHT WT66;
TABLES HIGH_LOW*CCC6_1C/CHISQ MEASURE ALPHA=0.01;
RUN;
************************/
data data;
set 'e:\temp4\data';
keep CCC6_1C WT66 PAC6DEE PAC6DLEI PAC6DFR PAC6DFD PAC6DPAI;
IF PAC6DEE<=99.6;
RUN;
/******
DATA TEST;
SET DATA;
IF LFC6GI13=1 OR LFC6GI13=11;
IF LFC6GI13=11 THEN LFC6GI13=0;
PROC FREQ DATA=TEST;
WEIGHT WT66;
TABLES LFC6GI13*CCC6_1C/CHISQ MEASURE ALPHA=0.01;
RUN;
******/
DATA TEST;
SET DATA;
IF PAC6DEE=99.6 THEN CH_AD=0;ELSE CH_AD=1;
PROC FREQ DATA=TEST;
WEIGHT WT66;
TABLES CH_AD*CCC6_1C/CHISQ MEASURE ALPHA=0.01;
RUN;

DATA TEST;
SET DATA;
```

```
IF PAC6DFR<=3;
IF PAC6DFR>1 THEN PAC6DFR=4;
PROC FREQ DATA=TEST;
WEIGHT WT66;
TABLES PAC6DFR*CCC6_1C/CHISQ MEASURE ALPHA=0.01;
RUN;

DATA TEST;
SET DATA;
IF PAC6DPAI<=3;
IF PAC6DPAI>1 THEN PAC6DPAI=4;
PROC FREQ DATA=TEST;
WEIGHT WT66;
TABLES PAC6DPAI*CCC6_1C/CHISQ MEASURES ALPHA=0.01;
RUN;

DATA TEST;
SET DATA;
IF PAC6DFD<=2;
PROC FREQ DATA=TEST;
WEIGHT WT66;
TABLES PAC6DFD*CCC6_1C/CHISQ MEASURE ALPHA=0.01;
RUN;
************************/
data data;
set 'e:\temp4\data';
keep CCC6_1C WT66 SDC6GCB SDC6GRAC DHC6GMAR;
RUN;

DATA TEST;
SET DATA;
IF     SDC6GCB=     1      THEN   SDC6G1        =0;    ELSE   SDC6G1        =1;
IF     SDC6GCB=     2      THEN   SDC6G2        =0;    ELSE   SDC6G2        =1;
IF     SDC6GCB=     3      THEN   SDC6G3        =0;    ELSE   SDC6G3        =1;
IF     SDC6GCB=     4      THEN   SDC6G4        =0;    ELSE   SDC6G4        =1;
PROC FREQ DATA=TEST;
WEIGHT WT66;
TABLES (SDC6G1-SDC6G4)*CCC6_1C/CHISQ MEASURE ALPHA=0.01;
RUN;

DATA TEST;
SET DATA;
if SDC6GRAC<=2;
PROC FREQ DATA=TEST;
WEIGHT WT66;
TABLES (SDC6grac)*CCC6_1C/CHISQ MEASURE ALPHA=0.01;
RUN;

DATA TEST;
SET DATA;
if DHC6GMAR<=3;
if DHC6GMAR ne 2 then DHC6GMAR=4;
PROC FREQ DATA=TEST;
WEIGHT WT66;
TABLES (DHC6GMAR)*CCC6_1C/CHISQ MEASURE ALPHA=0.01;
RUN;
DATA TEST;
```

```
SET DATA;
if DHC6GMAR=1 or DHC6GMAR=3;
if DHC6GMAR=1 then DHC6GMAR=4;
PROC FREQ DATA=TEST;
WEIGHT WT66;
TABLES (DHC6GMAR)*CCC6_1C/CHISQ MEASURE ALPHA=0.01;
RUN;
************************/
data data;
set 'e:\temp4\data';
keep CCC6_1C WT66 smc6_4 smc6_3 dhc6gage;
if smc6_3<=996;
RUN;

data test;
set data;
if smc6_3<996 then smoke=1; else smoke=2;
proc freq data=test;
weight wt66;
tables smoke*ccc6_1c/chisq measures alpha=0.01;
run;

data test;
set data;
if smc6_3 ne 996;
if smc6_3<20 then smoke=1; else smoke=2;
proc freq data=test;
weight wt66;
tables smoke*ccc6_1c/chisq measures alpha=0.01;
run;

data test;
set data;
if smc6_4 ne 996;
if smc6_4<10 then smoke=1; else smoke=2;
proc freq data=test;
weight wt66;
tables smoke*ccc6_1c/chisq measures alpha=0.01;
run;

data test;
set data;
if smc6_4 ne 996 and dhc6gage<=6;
if smc6_4>10 then smoke=1; else smoke=2;
proc freq data=test;
weight wt66;
tables smoke*ccc6_1c/chisq measures alpha=0.01;
run;

data test;
set data;
if smc6_4 ne 996 and dhc6gage>6;
if smc6_4>10 then smoke=1; else smoke=2;
proc freq data=test;
weight wt66;
tables smoke*ccc6_1c/chisq measures alpha=0.01;
run;
```

```
***********************/
data data;
set 'e:\temp4\data';
keep CCC6_1C WT66 dhc6_own DHC6GBED;
RUN;

data test;
set data;
if dhc6_own<=2 ;
IF dhc6_own=1 THEN dhc6_own=3;
proc freq data=test;
weight wt66;
tables dhc6_own*ccc6_1c/chisq measures alpha=0.01;
run;
***********************/
data data;
set 'e:\temp4\data';
keep CCC6_1C WT66 DHC6_SEX CCC6_1H HWC6GSW DHC6GAGE ;
RUN;
data test;
set data;
*if dhc6_own<=2 ;
IF dhc6_SEX=1 THEN dhc=4;ELSE DHC=3;
proc freq data=TEST;
weight wt66;
tables DHC*ccc6_1c/chisq measures alpha=0.01;
run;

data test;
set data;
if dhc6GAGE<=4 ;
*IF dhc6_SEX=1 THEN dhc=4;*ELSE DHC=3;
proc freq data=TEST;
weight wt66;
tables DHC6_SEX*ccc6_1c/chisq measures alpha=0.01;
run;

data test;
set data;
if dhc6GAGE>4 ;
IF dhc6_SEX=1 THEN dhc=4;ELSE DHC=3;
proc freq data=TEST;
weight wt66;
tables DHC*ccc6_1c/chisq measures alpha=0.01;
run;

data test;
set data;
if HWC6GSW<=4 ;
IF HWC6GSW<=3 THEN DHC=1;ELSE DHC=0;
proc freq data=TEST;
weight wt66;
tables DHC*ccc6_1c/chisq measures alpha=0.01;
run;
```

APPENDIX G

Programs used to fit loglinear regression model.

```
data loglin;
set 'e:\temp4\data' ;
if  0<=PRC6_CUR<=59
and GE36GURB<=6
AND HWC6G3KG<=138
AND HWC6GBMI<=99.6
AND HWC6GSW<=6
AND HWC6GHT<=64
AND HWS6_1<=6
AND HWS6_5<=6
AND LFC6GO21<=96
AND LFC6GI13<=96
AND PAC6DEE<=99.6
AND PAC6DLEI<=6
AND PAC6DFM<=996
AND PAC6DFR<=6
AND PAC6DFD<=6
AND PAC6DPAI<=9
AND SDC6GCB<=4
AND SDC6GRAC<=2
AND SMC6_3<=996
AND SMC6_4<=996
AND DGC6_1G<=6
AND CCC6_1H<=2
AND DGK6_1<=6
AND DHC6_OWN<=2
AND DHC6GBED<=5
AND DHC6GMAR<=3
AND DHC6GECF<=99
AND DHC6DLVG<=10;
run;
data new;
set loglin;
*keep CCC6_1C WT66 PRC6_CUR LFC6GI13 PAC6DPAI
*     SMC6_3 DHC6_OWN DHC6_SEX DHC6GAGE;
IF PRC6_CUR=12 OR PRC6_CUR=24 OR PRC6_CUR=35
   THEN PRC6_CUR=0; ELSE PRC6_CUR=1;
*IF 9<=LFC6GI13<=11 THEN LFC6GI13=0;* ELSE LFC6GI13=1;
IF 9<=LFC6GI13<=11 OR LFC6GI13=6 OR LFC6GI13=8 OR 12<=LFC6GI13<=13
   THEN LFC6GI13=1;
     ELSE IF LFC6GI13=96 THEN LFC6GI13=2;
       ELSE LFC6GI13=0;

IF  PAC6DPAI=1 THEN PAC6DPAI=0;
   ELSE IF PAC6DPAI>1 AND PAC6DPAI NE 6 THEN PAC6DPAI=1;
      ELSE PAC6DPAI=2;
IF SMC6_3=996 THEN SMC6_3=0; ELSE SMC6_3=1;
IF SDC6GCB>1 THEN SDC6GCB=0; ELSE SDC6GCB=1;
```

```
IF DHC6GBED<=2 THEN DHC6GBED=1; ELSE DHC6GBED=0;

RUN;
PROC CATMOD DATA=NEW;
WEIGHT WT66;
MODEL CCC6_1C*PRC6_CUR* LFC6GI13 *PAC6DPAI *DHC6_SEX
      *SMC6_3 *DHC6_OWN =_RESPONSE_
      /NOPROFILE NORESPONSE NOITER NOPARM;
LOGLIN CCC6_1C|PRC6_CUR|  LFC6GI13 |PAC6DPAI |DHC6_SEX
        |SMC6_3 |DHC6_OWN @5;
RUN;
```

APPENDIX H

Programs used to fit logistic regression model for all respondents.

```
data logistic;
set 'e:\temp4\data' ;
if  0<=PRC6_CUR<=59
and GE36GURB<=6
AND HWC6G3KG<=138
AND HWC6GBMI<=99.6
AND HWC6GSW<=6
AND HWC6GHT<=64
AND HWS6_1<=6
AND HWS6_5<=6
AND LFC6GO21<=96
AND LFC6GI13<=96
AND PAC6DEE<=99.6
AND PAC6DLEI<=6
AND PAC6DFM<=996
AND PAC6DFR<=6
AND PAC6DFD<=6
AND PAC6DPAI<=9
AND SDC6GCB<=4
AND SDC6GRAC<=2
AND SMC6_3<=996
AND SMC6_4<=996
AND DGC6_1G<=6
AND CCC6_1H<=2
AND DGK6_1<=6
AND DHC6_OWN<=2
AND DHC6GBED<=5
AND DHC6GMAR<=3
AND DHC6GECF<=99
AND DHC6DLVG<=10;
*if PRC6_CUR=10 THEN NFLD=0; *ELSE NFLD=0;
IF PRC6_CUR=11 THEN PEI=1; ELSE PEI=0;
IF PRC6_CUR=12 THEN NS=1; ELSE NS =0;
IF PRC6_CUR=13 THEN NB=1; ELSE NB=0;
IF PRC6_CUR=24 THEN QUE=1; ELSE QUE=0;
IF PRC6_CUR=35 THEN ONT=1; ELSE ONT=0;
IF PRC6_CUR=46 THEN MB=1; ELSE MB=0;
IF PRC6_CUR=47 THEN SASK=1; ELSE SASK=0;
IF PRC6_CUR=48 THEN AB=1; ELSE AB=0;
IF PRC6_CUR=59 THEN BC=1; ELSE BC=0;

IF 0<=HWC6GBMI<=57.9    THEN  HWC6GB1     =HWC6GBMI;  ELSE  HWC6GB1=0;

IF HWC6GSW=1      THEN  HWC6G1      =1;  ELSE  HWC6G1      =0;
IF HWC6GSW=2      THEN  HWC6G2      =1;  ELSE  HWC6G2      =0;
IF HWC6GSW=3      THEN  HWC6G3      =1;  ELSE  HWC6G3      =0;
IF HWC6GSW=4      THEN  HWC6G4      =1;  ELSE  HWC6G4      =0;
IF    HWS6_1=1    THEN  HWS1  =1;   ELSE  HWS1  =0;
IF    HWS6_1=2    THEN  HWS2  =1;   ELSE  HWS2  =0;
```

```
IF     HWS6_1=3     THEN  HWS3   =1;    ELSE  HWS3   =0;

* IF   0<=HWS6_2KG<=185 THEN   HWS6KG1      =HWS6_2KG;   * ELSE
       HWS6KG1      =0;

IF     HWS6_5=      1      THEN  HWS61 =1;    ELSE  HWS61 =0;
IF     HWS6_5=      2      THEN  HWS62 =1;    ELSE  HWS62 =0;


IF     LFC6GI13=    1      THEN  LFC1  =1;    ELSE  LFC1    =0;
IF     LFC6GI13=    2      THEN  LFC2  =1;    ELSE  LFC2  =0;
IF     LFC6GI13=    3      THEN  LFC3  =1;    ELSE  LFC3  =0;
IF     LFC6GI13=    4      THEN  LFC4  =1;    ELSE  LFC4  =0;
IF     LFC6GI13=    5      THEN  LFC5  =1;    ELSE  LFC5  =0;
IF     LFC6GI13=    6      THEN  LFC6  =1;    ELSE  LFC6  =0;
IF     LFC6GI13=    7      THEN  LFC7  =1;    ELSE  LFC7  =0;
IF     LFC6GI13=    8      THEN  LFC8  =1;    ELSE  LFC8  =0;
IF     LFC6GI13=    9      THEN  LFC9  =1;    ELSE  LFC9  =0;
IF     LFC6GI13=    10     THEN  LFC10 =1;    ELSE  LFC10 =0;
IF     LFC6GI13=    11     THEN  LFC11 =1;    ELSE  LFC11 =0;
IF     LFC6GI13=    12     THEN  LFC12 =1;    ELSE  LFC12 =0;
IF     LFC6GI13=    13     THEN  LFC13 =1;    ELSE  LFC13 =0;

IF     PAC6DLEI=1   THEN   PAC6DL1      =1;    ELSE  PAC6DL1      =0;
IF     PAC6DLEI=2   THEN   PAC6DL2      =1;    ELSE  PAC6DL2      =0;

IF     0<PAC6DFM<=255     THEN  PAC6DFM1     =PAC6DFM;
          ELSE if PAC6DFM=0 then PAC6DFM1=0.01;  ELSE PAC6DFM1=0;

IF     PAC6DPAI=    1      THEN  PAC6DPA1     =1;    ELSE  PAC6DPA1     =0;
IF     PAC6DPAI=    2      THEN  PAC6DPA2     =1;    ELSE  PAC6DPA2     =0;
IF     PAC6DPAI=    3      THEN  PAC6DPA3     =1;    ELSE  PAC6DPA3     =0;

IF     SDC6GCB=     1      THEN  SDC6G1       =1;    ELSE  SDC6G1       =0;
IF     SDC6GCB=     2      THEN  SDC6G2       =1;    ELSE  SDC6G2       =0;
IF     SDC6GCB=     3      THEN  SDC6G3       =1;    ELSE  SDC6G3       =0;
*IF    SDC6GCB=     4      THEN  SDC6G4       =0;*   ELSE  SDC6G4       =0;

IF     SDC6GRAC=    1      THEN  SDC6GR1      =1;    ELSE  SDC6GR1      =0;
*IF    SDC6GRAC=    2      THEN  SDC6GR2      =0;    *ELSE       SDC6GR2=0;

IF     SMC6_3=996          THEN  SMC6_31      =0;    ELSE  SMC6_31      =100;
IF     0<=SMC6_3<=85 THEN        SMC6_32      =SMC6_3;     ELSE  SMC6_32=0;

IF     0<=SMC6_4<=99 THEN        SMC6_41=SMC6_4;    ELSE SMC6_41      =0;

*IF    DHC6_OWN=    1      THEN  DHC6_O1      =0;    *ELSE       DHC6_O1=0;
IF     DHC6_OWN=    2      THEN  DHC6_O2      =1;    ELSE  DHC6_O2=0;

*IF    DHC6GBED=    0      THEN  DHC6GB1      =1;    *ELSE       DHC6GB1=0;
*IF    DHC6GBED=    1      THEN  DHC6GB2      =1;    *ELSE       DHC6GB2=0;
*IF    DHC6GBED=    2      THEN  DHC6GB3      =1;    *ELSE       DHC6GB3=0;
*IF    DHC6GBED=    3      THEN  DHC6GB4      =1;    *ELSE       DHC6GB4=0;
*IF    DHC6GBED=    4      THEN  DHC6GB5      =1;    *ELSE       DHC6GB5=0;
*IF    DHC6GBED=    5      THEN  DHC6GB6      =1;    *ELSE       DHC6GB6=0;
IF     0<=DHC6GBED<=5      THEN  DHC6GB1      =DHC6GBED;  ELSE  DHC6GB1=0;
```

```
*IF   DHC6_SEX=   1     THEN  DHC6_S1     =1;*  ELSE  DHC6_S1     =0;
IF    DHC6_SEX=   2     THEN  DHC6_S2     =1;   ELSE  DHC6_S2     =0;

*IF   DHC6GMAR=   1     THEN  DHC6GM1     =1;   *ELSE DHC6GM1=0;
IF    DHC6GMAR=   2     THEN  DHC6GM2     =1;   ELSE  DHC6GM2     =0;
IF    DHC6GMAR=   3     THEN  DHC6GM3     =1;   ELSE  DHC6GM3     =0;

*IF   DHC6GHSZ=   1     THEN  DHC6GH1     =1;   *ELSE       DHC6GH1=0;
*IF   DHC6GHSZ=   2     THEN  DHC6GH2     =1;   *ELSE       DHC6GH2=0;
*IF   DHC6GHSZ=   3     THEN  DHC6GH3     =1;   *ELSE       DHC6GH3=0;
*IF   DHC6GHSZ=   4     THEN  DHC6GH4     =1;   *ELSE       DHC6GH4=0;
*IF   DHC6GHSZ=   5     THEN  DHC6GH5     =1;   *ELSE       DHC6GH5=0;

*IF   0<=DHC6GHSZ<=5    THEN  DHC6GH      =1;   *ELSE       DHC6GH=0;

IF    DHC6GLE5=   1     THEN  DHC6GLE1    =1;   ELSE  DHC6GLE1    =0;
*IF   DHC6GLE5=   2     THEN  DHC6GLE2    =1;*  ELSE  DHC6GLE2    =0;

IF    DHC6G611=   1     THEN  DHC6G61     =1;   ELSE  DHC6G61     =0;
*IF   DHC6G611=   2     THEN  DHC6G62     =1;   *ELSE DHC6G62     =0;

IF    DHC6GECF=   1     THEN  DHC6GE1     =1;   ELSE  DHC6GE1     =0;
IF    DHC6GECF=   2     THEN  DHC6GE2     =1;   ELSE  DHC6GE2     =0;
IF    DHC6GECF=   3     THEN  DHC6GE3     =1;   ELSE  DHC6GE3     =0;
IF    DHC6GECF=   4     THEN  DHC6GE4     =1;   ELSE  DHC6GE4     =0;
IF    DHC6GECF=   5     THEN  DHC6GE5     =1;   ELSE  DHC6GE5     =0;
IF    DHC6GECF=   6     THEN  DHC6GE6     =1;   ELSE  DHC6GE6     =0;
/*IF  DHC6GECF=   7     THEN  DHC6GE7     =1;   ELSE  DHC6GE7     =0;*/
IF    DHC6GECF=   8     THEN  DHC6GE8     =1;   ELSE  DHC6GE8     =0;
IF    DHC6GECF=   9     THEN  DHC6GE9     =1;   ELSE  DHC6GE9     =0;

IF    DHC6DLVG=   1     THEN  DHC6DL1     =1;   ELSE  DHC6DL1     =0;
IF    DHC6DLVG=   2     THEN  DHC6DL2     =1;   ELSE  DHC6DL2     =0;
IF    DHC6DLVG=   3     THEN  DHC6DL3     =1;   ELSE  DHC6DL3     =0;
IF    DHC6DLVG=   4     THEN  DHC6DL4     =1;   ELSE  DHC6DL4     =0;
IF    DHC6DLVG=   5     THEN  DHC6DL5     =1;   ELSE  DHC6DL5     =0;
IF    DHC6DLVG=   6     THEN  DHC6DL6     =1;   ELSE  DHC6DL6     =0;
/*IF  DHC6DLVG=   7     THEN  DHC6DL7     =1;   ELSE  DHC6DL7     =0;*/
IF    DHC6DLVG=   8     THEN  DHC6DL8     =1;   ELSE  DHC6DL8     =0;
IF    DHC6DLVG=   9     THEN  DHC6DL9     =1;   ELSE  DHC6DL9     =0;
IF    DHC6DLVG=   10    THEN  DHC6DL10    =1;   ELSE  DHC6DL10    =0;
run;

proc logistic data =lgtdata DESCENDING;
weight WT66;
model CCC6_1C=
NFLD PEI  NB QUE ONT MB SASK AB BC
GE36GU1-GE36GU2 HWC6GB1 HWC6G1-HWC6G4
HWS61-HWS62 LFC1-LFC13 PAC6DFM1
PAC6DPA1-PAC6DPA3 SDC6G1-SDC6G3
SMC6_31 SMC6_32 SMC6_41
DHC6_O2 DHC6GB1 DHC6_S2 DHC6GM2
DHC6GM3 DHC6G61 DHC6GE1-DHC6GE6
DHC6GE8-DHC6GE9 DHC6DL2-DHC6DL6
DHC6DL8-DHC6DL9 HWC6GHT DHC6GAGE
/lackfit SELECTION=STEPWISE link=logit;
run;
```

```
/*
link=logit
link=probit
link=cloglog

proc logistic data =lgtdata DESCENDING;
weight WT66;
model CCC6_1C=
NFLD   NB    AB     GE36GU1  HWC6G1
HWC6G2       HWC6G3      LFC1 LFC3 LFC6
LFC9   LFC10  LFC13 SMC6_31 SMC6_41
LFC11  LFC12        PAC6DPA1 PAC6DPA2
PAC6DPA3  SDC6G1   SDC6G2      SDC6G3
DHC6_O2        DHC6_S2    DHC6GM2 DHC6GM3
DHC6GE1        DHC6GE4    DHC6GE5 DHC6GE6
DHC6DL5   DHC6DL6    DHC6GAGE
/lackfit link=logit outroc=roc1 roceps=0;
output out=outp p=p;
run;
proc plot data=roc1;
plot (_SENSIT_) *(_1MSPEC_);
run;
```

In Splus:

```
>plot(c(-0.0,1),c(-0.0,1),type='n', xlab='1-Specificity',
ylab='Sensitivity')
>lines(roc1$.1MSPEC,roc1$.SENSIT.)
>lines(c(0,1),c(0,1))
```

APPENDIX I

Programs used for stepwise logistic regression and cross validation for infants, children and adults.

```
data lgtdata1;
set 'e:\temp4\data';

if  PRC6_CUR<=60
and GE36GURB<=6
AND HWC6G3KG<=138
AND HWC6GBMI<=99.6
AND HWC6GSW<=6
AND HWC6GHT<=64
AND HWS6_1<=6
AND HWS6_5<=6
AND LFC6GO21<=96
AND LFC6GI13<=96
AND PAC6DEE<=99.6
AND PAC6DLEI<=6
AND PAC6DFM<=996
AND PAC6DFR<=6
AND PAC6DFD<=6
AND PAC6DPAI<=9
AND SDC6GCB<=4
AND SDC6GRAC<=2
AND SMC6_3<=996
AND SMC6_4<=996
AND DGC6_1G<=6
AND CCC6_1H<=2
AND DGK6_1<=6
AND DHC6_OWN<=2
AND DHC6GBED<=5
AND DHC6GMAR<=3
AND DHC6GECF<=99
AND DHC6DLVG<=10;

if PRC6_CUR=10 THEN NFLD=1; ELSE NFLD=0;
IF PRC6_CUR=11 THEN PEI=1; ELSE PEI=0;
/*IF PRC6_CUR=12 THEN NS=1; ELSE NS =0; */
IF PRC6_CUR=13 THEN NB=1; ELSE NB=0;
IF PRC6_CUR=24 THEN QUE=1; ELSE QUE=0;
IF PRC6_CUR=35 THEN ONT=1; ELSE ONT=0;
IF PRC6_CUR=46 THEN MB=1; ELSE MB=0;
IF PRC6_CUR=47 THEN SASK=1; ELSE SASK=0;
IF PRC6_CUR=48 THEN AB=1; ELSE AB=0;
IF PRC6_CUR=59 THEN BC=1; ELSE BC=0;

IF    GE36GURB=   1     THEN  GE36GU1     =1;    ELSE  GE36GU1     =0;
IF    GE36GURB=   2     THEN  GE36GU2     =1;    ELSE  GE36GU2     =0;

IF    GE36GCMA=   462   THEN  GE36GC1     =1;    ELSE  GE36GC1     =0;
IF    GE36GCMA=   933   THEN  GE36GC2     =1;    ELSE  GE36GC2     =0;
```

```
IF    GE36GHLR=   461    THEN  GE36GH1     =1;   ELSE  GE36GH1     =0;
IF    GE36GHLR=   462    THEN  GE36GH2     =1;   ELSE  GE36GH2     =0;
IF    GE36GHLR=   463    THEN  GE36GH3     =1;   ELSE  GE36GH3     =0;
IF    GE36GHLR=   464    THEN  GE36GH4     =1;   ELSE  GE36GH4     =0;
IF    GE36GHLR=   465    THEN  GE36GH5     =1;   ELSE  GE36GH5     =0;
IF    GE36GHLR=   481    THEN  GE36GH6     =1;   ELSE  GE36GH6     =0;
IF    GE36GHLR=   482    THEN  GE36GH7     =1;   ELSE  GE36GH7     =0;
IF    GE36GHLR=   483    THEN  GE36GH8     =1;   ELSE  GE36GH8     =0;
IF    GE36GHLR=   484    THEN  GE36GH9     =1;   ELSE  GE36GH9     =0;
IF    GE36GHLR=   485    THEN  GE36GH10    =1;   ELSE  GE36GH10    =0;
IF    GE36GHLR=   3511   THEN  GE36GH11    =1;   ELSE  GE36GH11    =0;
IF    GE36GHLR=   3512   THEN  GE36GH12    =1;   ELSE  GE36GH12    =0;
IF    GE36GHLR=   3521   THEN  GE36GH13    =1;   ELSE  GE36GH13    =0;
IF    GE36GHLR=   3522   THEN  GE36GH14    =1;   ELSE  GE36GH14    =0;
IF    GE36GHLR=   3523   THEN  GE36GH15    =1;   ELSE  GE36GH15    =0;
IF    GE36GHLR=   3524   THEN  GE36GH16    =1;   ELSE  GE36GH16    =0;
IF    GE36GHLR=   3531   THEN  GE36GH17    =1;   ELSE  GE36GH17    =0;
IF    GE36GHLR=   3532   THEN  GE36GH18    =1;   ELSE  GE36GH18    =0;
IF    GE36GHLR=   3533   THEN  GE36GH19    =1;   ELSE  GE36GH19    =0;
IF    GE36GHLR=   3534   THEN  GE36GH20    =1;   ELSE  GE36GH20    =0;
IF    GE36GHLR=   3541   THEN  GE36GH21    =1;   ELSE  GE36GH21    =0;
IF    GE36GHLR=   3542   THEN  GE36GH22    =1;   ELSE  GE36GH22    =0;
IF    GE36GHLR=   3543   THEN  GE36GH23    =1;   ELSE  GE36GH23    =0;
IF    GE36GHLR=   3551   THEN  GE36GH24    =1;   ELSE  GE36GH24    =0;
IF    GE36GHLR=   3552   THEN  GE36GH25    =1;   ELSE  GE36GH25    =0;
IF    GE36GHLR=   3561   THEN  GE36GH26    =1;   ELSE  GE36GH26    =0;

IF    GE36GHRO=   461    THEN  GE36G1      =1;   ELSE  GE36G1      =0;
IF    GE36GHRO=   462    THEN  GE36G2      =1;   ELSE  GE36G2      =0;
IF    GE36GHRO=   463    THEN  GE36G3      =1;   ELSE  GE36G3      =0;
IF    GE36GHRO=   464    THEN  GE36G4      =1;   ELSE  GE36G4      =0;
IF    GE36GHRO=   465    THEN  GE36G5      =1;   ELSE  GE36G5      =0;
IF    GE36GHRO=   481    THEN  GE36G6      =1;   ELSE  GE36G6      =0;
IF    GE36GHRO=   482    THEN  GE36G7      =1;   ELSE  GE36G7      =0;
IF    GE36GHRO=   483    THEN  GE36G8      =1;   ELSE  GE36G8      =0;
IF    GE36GHRO=   484    THEN  GE36G9      =1;   ELSE  GE36G9      =0;
IF    GE36GHRO=   485    THEN  GE36G10     =1;   ELSE  GE36G10     =0;
IF    GE36GHRO=   3511   THEN  GE36G11     =1;   ELSE  GE36G11     =0;
IF    GE36GHRO=   3512   THEN  GE36G12     =1;   ELSE  GE36G12     =0;
IF    GE36GHRO=   3513   THEN  GE36G13     =1;   ELSE  GE36G13     =0;
IF    GE36GHRO=   3521   THEN  GE36G14     =1;   ELSE  GE36G14     =0;
IF    GE36GHRO=   3522   THEN  GE36G15     =1;   ELSE  GE36G15     =0;
IF    GE36GHRO=   3523   THEN  GE36G16     =1;   ELSE  GE36G16     =0;
IF    GE36GHRO=   3524   THEN  GE36G17     =1;   ELSE  GE36G17     =0;
IF    GE36GHRO=   3525   THEN  GE36G18     =1;   ELSE  GE36G18     =0;
IF    GE36GHRO=   3526   THEN  GE36G19     =1;   ELSE  GE36G19     =0;
IF    GE36GHRO=   3527   THEN  GE36G20     =1;   ELSE  GE36G20     =0;
IF    GE36GHRO=   3531   THEN  GE36G21     =1;   ELSE  GE36G21     =0;
IF    GE36GHRO=   3532   THEN  GE36G22     =1;   ELSE  GE36G22     =0;
IF    GE36GHRO=   3533   THEN  GE36G23     =1;   ELSE  GE36G23     =0;
IF    GE36GHRO=   3534   THEN  GE36G24     =1;   ELSE  GE36G24     =0;
IF    GE36GHRO=   3536   THEN  GE36G25     =1;   ELSE  GE36G25     =0;
IF    GE36GHRO=   3541   THEN  GE36G26     =1;   ELSE  GE36G26     =0;
IF    GE36GHRO=   3542   THEN  GE36G27     =1;   ELSE  GE36G27     =0;
IF    GE36GHRO=   3543   THEN  GE36G28     =1;   ELSE  GE36G28     =0;
IF    GE36GHRO=   3544   THEN  GE36G29     =1;   ELSE  GE36G29     =0;
```

```
IF    GE36GHRO=   3551  THEN  GE36G30      =1;   ELSE  GE36G30      =0;
IF    GE36GHRO=   3552  THEN  GE36G31      =1;   ELSE  GE36G31      =0;
IF    GE36GHRO=   3553  THEN  GE36G32      =1;   ELSE  GE36G32      =0;
IF    GE36GHRO=   3561  THEN  GE36G33      =1;   ELSE  GE36G33      =0;

IF    0<=HWC6GBMI<=57.9 THEN  HWC6GB1      =HWC6GBMI;  ELSE  HWC6GB1
      =0;

IF    HWC6GSW=    1     THEN  HWC6G1       =1;   ELSE  HWC6G1       =0;
IF    HWC6GSW=    2     THEN  HWC6G2       =1;   ELSE  HWC6G2       =0;
IF    HWC6GSW=    3     THEN  HWC6G3       =1;   ELSE  HWC6G3       =0;
IF    HWC6GSW=    4     THEN  HWC6G4       =1;   ELSE  HWC6G4       =0;

IF    HWS6_1=     1     THEN  HWS1  =1;   ELSE  HWS1  =0;
IF    HWS6_1=     2     THEN  HWS2  =1;   ELSE  HWS2  =0;
IF    HWS6_1=     3     THEN  HWS3  =1;   ELSE  HWS3  =0;

* IF  0<=HWS6_2KG<=185 THEN   HWS6KG1      =HWS6_2KG;  * ELSE
      HWS6KG1     =0;

IF    HWS6_5=     1     THEN  HWS61 =1;   ELSE  HWS61 =0;
IF    HWS6_5=     2     THEN  HWS62 =1;   ELSE  HWS62 =0;

IF    LFC6GO21=   1     THEN  LFC6G1       =1;   ELSE  LFC6G1       =0;
IF    LFC6GO21=   2     THEN  LFC6G2       =1;   ELSE  LFC6G2       =0;
IF    LFC6GO21=   3     THEN  LFC6G3       =1;   ELSE  LFC6G3       =0;
IF    LFC6GO21=   4     THEN  LFC6G4       =1;   ELSE  LFC6G4       =0;
IF    LFC6GO21=   5     THEN  LFC6G5       =1;   ELSE  LFC6G5       =0;
IF    LFC6GO21=   6     THEN  LFC6G6       =1;   ELSE  LFC6G6       =0;
IF    LFC6GO21=   7     THEN  LFC6G7       =1;   ELSE  LFC6G7       =0;
IF    LFC6GO21=   8     THEN  LFC6G8       =1;   ELSE  LFC6G8       =0;
IF    LFC6GO21=   9     THEN  LFC6G9       =1;   ELSE  LFC6G9       =0;
IF    LFC6GO21=   10    THEN  LFC6G10      =1;   ELSE  LFC6G10      =0;
IF    LFC6GO21=   11    THEN  LFC6G11      =1;   ELSE  LFC6G11      =0;
IF    LFC6GO21=   12    THEN  LFC6G12      =1;   ELSE  LFC6G12      =0;
IF    LFC6GO21=   13    THEN  LFC6G13      =1;   ELSE  LFC6G13      =0;
IF    LFC6GO21=   14    THEN  LFC6G14      =1;   ELSE  LFC6G14      =0;
IF    LFC6GO21=   15    THEN  LFC6G15      =1;   ELSE  LFC6G15      =0;
IF    LFC6GO21=   16    THEN  LFC6G16      =1;   ELSE  LFC6G16      =0;
IF    LFC6GO21=   17    THEN  LFC6G17      =1;   ELSE  LFC6G17      =0;
IF    LFC6GO21=   18    THEN  LFC6G18      =1;   ELSE  LFC6G18      =0;
IF    LFC6GO21=   19    THEN  LFC6G19      =1;   ELSE  LFC6G19      =0;
IF    LFC6GO21=   20    THEN  LFC6G20      =1;   ELSE  LFC6G20      =0;
IF    LFC6GO21=   21    THEN  LFC6G21      =1;   ELSE  LFC6G21      =0;

IF    LFC6GI13=   96    THEN  LFC1  =1;   ELSE  LFC1    =0;
IF    LFC6GI13=   2     THEN  LFC2  =1;   ELSE  LFC2  =0;
IF    LFC6GI13=   3     THEN  LFC3  =1;   ELSE  LFC3  =0;
IF    LFC6GI13=   4     THEN  LFC4  =1;   ELSE  LFC4  =0;
IF    LFC6GI13=   5     THEN  LFC5  =1;   ELSE  LFC5  =0;
IF    LFC6GI13=   6     THEN  LFC6  =1;   ELSE  LFC6  =0;
IF    LFC6GI13=   7     THEN  LFC7  =1;   ELSE  LFC7  =0;
IF    LFC6GI13=   8     THEN  LFC8  =1;   ELSE  LFC8  =0;
IF    LFC6GI13=   9     THEN  LFC9  =1;   ELSE  LFC9  =0;
IF    LFC6GI13=   10    THEN  LFC10 =1;   ELSE  LFC10 =0;
IF    LFC6GI13=   11    THEN  LFC11 =1;   ELSE  LFC11 =0;
IF    LFC6GI13=   12    THEN  LFC12 =1;   ELSE  LFC12 =0;
```

```
IF     LFC6GI13=   13     THEN  LFC13 =1;    ELSE  LFC13 =0;


IF     0<=PAC6DEE<=34.4  THEN  PAC6DE1      =PAC6DEE;   ELSE  PAC6DE1
       =0;

IF     PAC6DLEI=1  THEN  PAC6DL1      =1;    ELSE  PAC6DL1      =0;
IF     PAC6DLEI=2  THEN  PAC6DL2      =1;    ELSE  PAC6DL2      =0;

IF     0<PAC6DFM<=255    THEN  PAC6DFM1     =PAC6DFM;
     ELSE if PAC6DFM=0 THEN PAC6DFM1=0.01;
         ELSE PAC6DFM1=0;

IF     PAC6DFR=     1      THEN  PAC6DFR1     =1;    ELSE  PAC6DFR1     =0;
IF     PAC6DFR=     2      THEN  PAC6DFR2     =1;    ELSE  PAC6DFR2     =0;
IF     PAC6DFR=     3      THEN  PAC6DFR3     =1;    ELSE  PAC6DFR3     =0;

IF     PAC6DFD=     1      THEN  PAC6DFD1     =1;    ELSE  PAC6DFD1     =0;
IF     PAC6DFD=     2      THEN  PAC6DFD2     =1;    ELSE  PAC6DFD2     =0;

IF     PAC6DPAI=    1      THEN  PAC6DPA1     =1;    ELSE  PAC6DPA1     =0;
IF     PAC6DPAI=    2      THEN  PAC6DPA2     =1;    ELSE  PAC6DPA2     =0;
IF     PAC6DPAI=    3      THEN  PAC6DPA3     =1;    ELSE  PAC6DPA3     =0;

IF     SDC6GCB=     1      THEN  SDC6G1       =1;    ELSE  SDC6G1       =0;
IF     SDC6GCB=     2      THEN  SDC6G2       =1;    ELSE  SDC6G2       =0;
IF     SDC6GCB=     3      THEN  SDC6G3       =1;    ELSE  SDC6G3       =0;
/* IF SDC6GCB=     4      THEN  SDC6G4       =1;    ELSE  SDC6G4       =0;
*/

IF     SDC6GRAC=    1      THEN  SDC6GR1      =1;    ELSE  SDC6GR1      =0;
/*IF   SDC6GRAC=    2      THEN  SDC6GR2      =1;    ELSE  SDC6GR2      =0;*/

IF     0<=SMC6_3<=85             THEN  SMC6_31      =SMC6_3;     ELSE
       SMC6_31      =100;
IF     SMC6_3=996          THEN  SMC6_32      =0;    ELSE  SMC6_32      =1;

IF     0<=SMC6_4<=99             THEN  SMC6_41      =SMC6_4;     ELSE
       SMC6_41      =0;

IF     DGC6_1G=     1      THEN  DGC6G1       =1;    ELSE  DGC6G1       =0;
IF     DGC6_1G=     2      THEN  DGC6G2       =1;    ELSE  DGC6G2       =0;

IF     CCC6_1H=     1      THEN  CCC6H1       =1;    ELSE  CCC6H1       =0;
/*IF   CCC6_1H=     2      THEN  CCC6H2       =1;    ELSE  CCC6H2       =0;
*/

IF     DGK6_1=      1      THEN  DGK61 =1;    ELSE  DGK61 =0;
IF     DGK6_1=      2      THEN  DGK62 =1;    ELSE  DGK62 =0;

/*IF   DHC6_OWN=    1      THEN  DHC6_O1      =1;    ELSE  DHC6_O1      =0;*/
IF     DHC6_OWN=    2      THEN  DHC6_O2      =1;    ELSE  DHC6_O2      =0;

*IF    DHC6GBED=    0      THEN  DHC6GB1      =1;    *ELSE        DHC6GB1
       =0;
*IF    DHC6GBED=    1      THEN  DHC6GB2      =1;    *ELSE        DHC6GB2
       =0;
```

```
*IF    DHC6GBED=    2     THEN  DHC6GB3      =1;    *ELSE        DHC6GB3
       =0;
*IF    DHC6GBED=    3     THEN  DHC6GB4      =1;    *ELSE        DHC6GB4
       =0;
*IF    DHC6GBED=    4     THEN  DHC6GB5      =1;    *ELSE        DHC6GB5
       =0;
*IF    DHC6GBED=    5     THEN  DHC6GB6      =1;    *ELSE        DHC6GB6
       =0;
IF     0<=DHC6GBED<=5     THEN  DHC6GB1      =DHC6GBED;  ELSE    DHC6GB1
       =0;

/*IF   DHC6_SEX=    1     THEN  DHC6_S1      =1;    ELSE  DHC6_S1      =0;*/
IF     DHC6_SEX=    2     THEN  DHC6_S2      =1;    ELSE  DHC6_S2      =0;

/*IF   DHC6GMAR=    1     THEN  DHC6GM1      =1;    ELSE  DHC6GM1      =0;*/
IF     DHC6GMAR=    2     THEN  DHC6GM2      =1;    ELSE  DHC6GM2      =0;
IF     DHC6GMAR=    3     THEN  DHC6GM3      =1;    ELSE  DHC6GM3      =0;

*IF    DHC6GHSZ=    1     THEN  DHC6GH1      =1;    *ELSE        DHC6GH1
       =0;
*IF    DHC6GHSZ=    2     THEN  DHC6GH2      =1;    *ELSE        DHC6GH2
       =0;
*IF    DHC6GHSZ=    3     THEN  DHC6GH3      =1;    *ELSE        DHC6GH3
       =0;
*IF    DHC6GHSZ=    4     THEN  DHC6GH4      =1;    *ELSE        DHC6GH4
       =0;
*IF    DHC6GHSZ=    5     THEN  DHC6GH5      =1;    *ELSE        DHC6GH5
       =0;
IF     0<=DHC6GHSZ<=5     THEN  DHC6GH       =1;    ELSE  DHC6GH=0;

IF     DHC6GLE5=    1     THEN  DHC6GLE1     =1;    ELSE  DHC6GLE1     =0;
/*IF   DHC6GLE5=    2     THEN  DHC6GLE2     =1;    ELSE  DHC6GLE2     =0;*/

IF     DHC6G611=    1     THEN  DHC6G61      =1;    ELSE  DHC6G61      =0;
/*IF   DHC6G611=    2     THEN  DHC6G62      =1;    ELSE  DHC6G62      =0;*/

IF     DHC6GECF=    1     THEN  DHC6GE1      =1;    ELSE  DHC6GE1      =0;
IF     DHC6GECF=    2     THEN  DHC6GE2      =1;    ELSE  DHC6GE2      =0;
IF     DHC6GECF=    3     THEN  DHC6GE3      =1;    ELSE  DHC6GE3      =0;
IF     DHC6GECF=    4     THEN  DHC6GE4      =1;    ELSE  DHC6GE4      =0;
IF     DHC6GECF=    5     THEN  DHC6GE5      =1;    ELSE  DHC6GE5      =0;
IF     DHC6GECF=    6     THEN  DHC6GE6      =1;    ELSE  DHC6GE6      =0;
/*IF   DHC6GECF=    7     THEN  DHC6GE7      =1;    ELSE  DHC6GE7      =0;*/
IF     DHC6GECF=    8     THEN  DHC6GE8      =1;    ELSE  DHC6GE8      =0;
IF     DHC6GECF=    9     THEN  DHC6GE9      =1;    ELSE  DHC6GE9      =0;

IF     DHC6DLVG=    1     THEN  DHC6DL1      =1;    ELSE  DHC6DL1      =0;
IF     DHC6DLVG=    2     THEN  DHC6DL2      =1;    ELSE  DHC6DL2      =0;
IF     DHC6DLVG=    3     THEN  DHC6DL3      =1;    ELSE  DHC6DL3      =0;
IF     DHC6DLVG=    4     THEN  DHC6DL4      =1;    ELSE  DHC6DL4      =0;
IF     DHC6DLVG=    5     THEN  DHC6DL5      =1;    ELSE  DHC6DL5      =0;
IF     DHC6DLVG=    6     THEN  DHC6DL6      =1;    ELSE  DHC6DL6      =0;
/*IF   DHC6DLVG=    7     THEN  DHC6DL7      =1;    ELSE  DHC6DL7      =0;*/
IF     DHC6DLVG=    8     THEN  DHC6DL8      =1;    ELSE  DHC6DL8      =0;
IF     DHC6DLVG=    9     THEN  DHC6DL9      =1;    ELSE  DHC6DL9      =0;
IF     DHC6DLVG=    10    THEN  DHC6DL10     =1;    ELSE  DHC6DL10     =0;
run;
```

```
data lgtdata;
set lgtdata1;
IF 6<DHC6GAGE;
run;

data lgtdata;
set lgtdata;
mul=rantbl(0,0.333333,0.333333,0.333334);
run;

data a;
set lgtdata;
if mul=1;
run;

data b;
set lgtdata;
if mul=2;
run;

data c;
set lgtdata;
if mul=3;
run;


**********

proc logistic data =a DESCENDING;
weight WT66;
model CCC6_1C=
NFLD PEI  NB QUE ONT MB SASK AB BC
GE36GU1-GE36GU2 HWC6GB1 HWC6G1-HWC6G4
HWS61-HWS62 LFC1-LFC13
PAC6DFM1
PAC6DPA1-PAC6DPA3 SDC6G1-SDC6G3
SMC6_31  SMC6_41
DHC6_O2 DHC6GB1 DHC6_S2 DHC6GM2-DHC6GM3
DHC6G61 DHC6GE1-DHC6GE6 DHC6GE8-DHC6GE9
DHC6DL2-DHC6DL6 DHC6DL8-DHC6DL9
HWC6GHT DHC6GAGE
/lackfit SELECTION=STEPWISE link=LOGIT;
run;

*************
proc logistic data =b DESCENDING;
weight WT66;
model CCC6_1C=
                        BC
                          GE36GU2
                          HWC6G1
                          HWC6G2
                          LFC1
                          LFC3
                          LFC7
                          LFC8
                          LFC13
```

```
                         PAC6DPA1
                         PAC6DPA2
                         SDC6G1
                         SDC6G2
                         DHC6_O2
                         DHC6_S2
                         DHC6GM2
                         DHC6G61
                         DHC6GE4
                         DHC6GE5
                         HWC6GHT
                         DHC6GAGE         ;
/ outroc=roc1 roceps=1e-5;
output out=outp p=p;
run;


**************
data out;
set c;
ee=EXP(         -2.5267
              -BC              *     0.1365
              -GE36GU2         *     0.2500
              -HWC6G1          *    0.00062
              -HWC6G2         *     0.2273
              +LFC1            *      0.3750
              -LFC3           *     0.2381
              -LFC7            *    0.4545
              -LFC8            *    0.2375
              +LFC13           *      0.2692
              +PAC6DPA1       *      0.0665
              +PAC6DPA2        *      0.0357
              +SDC6G1          *      0.9132
              +SDC6G2         *      0.8882
              +DHC6_O2         *      0.4687
              +DHC6_S2        *      0.2969
              -DHC6GM2         *    0.0262
              -DHC6G61        *    0.00820
              +DHC6GE4         *      0.5963
              -DHC6GE5        *    0.0189
              -HWC6GHT        *    0.00572
              -DHC6GAGE        *    0.0905
);
P=EE/(1+EE);
if ccc6_1c=1 then asthma=1; else asthma=0;
RUN;


data out;
SET OUT;
keep asthma p;
run;
proc sort data=out;
BY descending P;
run;

***************************
```

```
#In Splus:

roc_function(out)
{out<-as.matrix(out)
 # se<-0
 # sp1<-0
 # n<-0
 # m1<-0
 # m2<-0

 n<-dim(out)[1]
 m1<-sum(out[,2])
 m2<-n-m1
 s3<-rep(0,n)
 s4<-rep(0,n)
for (i in 1:n)
{
      s4[i]<-sum(out[1:i,2])
      s3[i]<-n-i-sum(out[i:n,2])

}
se<-s4/m1
sp1<-1-s3/m2
 lines(sp1,se)
}

#pp<-out[,1]
#p0<-out[1,1]
#for (i in 2:n)
#{
#     if (p0-pp[i]>=0.001)   p0<-pp[i]
#           else {
#            se[i]<-NA
#            sp1[i]<-NA
#           }
#}
#se<-se[!is.na(se)]
#sp1<-sp1[!is.na(sp1)]

plot(c(-0.0,1),c(-0.0,1),type='n', xlab='1-Specificity',
ylab='Sensitivity')
  lines(c(0,1),c(0,1))

roc(outic)
 lines(outib$.1MSPEC,outib$.SENSIT.,lty=3)

roc(outcc)
 lines(outcb$.1MSPEC,outcb$.SENSIT.,lty=3)

roc(outac)
 lines(outab$.1MSPEC,outab$.SENSIT.,lty=3)
```