# ECOLOGIC ANALYSIS OF LUNG AND STOMACH CANCER IN ONTARIO

BY

MICHELLE DENISE SHEBIB

A Research Paper

Submitted to the Department of Geography

in Fulfillment of the Requirements

of Geography 4C6

McMaster University

April 1991

ABSTRACT

Using maps, correlation and multiple regression, an ecologic analysis was performed to examine the geographic distribution of cancer incidence in Ontario with respect to selected ethnic, socio-economic and environmental characteristics for the 10 year period, 1976-1985. Two of the most common causes of cancer deaths, stomach and lung, were studied for each sex separately. The unit of analysis consisted of census divisions. The information used for the cancer were standardized incidence rates from the Ontario cancer Registry. The data for the ecologic variables was obtained from the 1981 Census of Canada. Two of the ecologic variables, education and income (low and median) were used to account for the effects of smoking.

Correlation co-efficients were significant for both sites of cancer for males and females for % urban and population density revealing the possibility of a positive relationship with cancer incidence and environmental characteristics. Ethnicity was strongly related to male and female stomach cancer.

Significant regression models were obtained for each of the cancer sites using a stepwise procedure with backward elimination. For each of the "best fit" equations, median income and education were included to control for smoking effects. Population density was significant in all equations at the 0.05 level. The percentage urban was significant for all except female stomach cancer. Manufacturing had a negative significant relationship for all cancer sites (male and female).

Also included in the study were descriptive statistics and cancer maps to determine the strongest cancer distributions in Ontario. For each site, northern Ontario contained the highest rates. In southern Ontario, urban areas such as Hamilton-Wentworth, and Toronto-York had high rates for all cancers (except Hamilton-Wentworth for male lung cancer).

ACKNOWLEDGEMENTS

I woulk like to than all of those people who helped with the creation of this research paper. First and foremost, I would like to thank my advisor Dr. S.M. Taylor for his constant guidance and assistance. The paper would not have been completed without his positive criticism and knowledge. I would also like to thank Dr. Liaw for his irreplaceable knowledge of computer methods (SAS and atlas graphics). His help allowed me to get over those constant hurdles. I would also like to than R. Hamilton for the instruction and use of his computer packages for mapping. Last but not least, I would like to thank all those in 303 who have made it a memorable year.

Michelle D. Shebib
April, 1991

TABLE OF CONTENTS

## LIST OF TABLES



## LIST OF FIGURES

# CHAPTER 1 : INTRODUCTION

In the field of medical geography, a major focus of study has been ecologic analyses of cancer. The primary reason for monitoring diseases (disease registries etc.) is to determine patterns that may lead to the development of avoidable risks. The purpose of this study is to determine the relationship between 4 of the most common cancer types in Ontario and environmental, socio-economic and ethnic variables for the 10 year period from 1975 - 1985. The cancer sites are lung and stomach cancer standardized by sex and age. The independent variables came from the census of Canada and included ethnicity (ie., British, Native, Italian, German, French), environment (ie. %urban, population density) and socio-economic (ie., income levels and education). The unit of analysis was the census divisions of counties in Ontario. In order to visualize distributions in Ontario, maps were generated for each of the dependent variables (cancer rates). This was followed by a more thorough statistical analysis using correlations and multiple regression techniques.

An purpose of ecologic studies is to examine relationships between living beings and the environment. In epidemiology, the ecologic approach is most often used to examine the association between a suspected exposure and a disease outcome for which data are summarized and compared

between groups rather than individuals (Reynolds, 1989). Ecologic studies are weak designs. They have their advantages and disadvantages (which will be discussed). They allow for an inexpensive meaningful analysis to be carried out on large populations; however, causal inferences cannot be made. Ecologic analysis mainly serve to generate hypotheses for future study.

The introduction section of the paper includes the problem at hand which is to examine the relationship between lung and stomach cancers in Ontario and selected environmentally related variables collected from the 1981 Census of Canada. The Literature Review section of the study provides background and supplementary information on epidemiological studies in the cancer field. They also give insight to previous studies that have been similar to this study. The background literature includes information on the relationships between the environment and human health as well as the relationships between the environment and the social determinants of cancer. The research methods are clearly explained in the methodology section of the paper. They include a presentation and description of maps, correlation and regression analyses. The conclusion section of the paper will give a summary of the findings as well as the consistency of those findings with other research. Finally, the directions for further research in the field are discussed.

# CHAPTER 2 : LITERATURE REVIEW

## 2.1 Introduction

One topic of research in the field of medical geography has been ecologic analyses. They attempt to make possible links between disease (cancer) and the environment. There are many problems with this type of approach. They are mainly hypothesis generating. The purpose of this chapter is to review the advantages and disadvantages of this method and previous studies and conclusions that have been made with respect to the environmental effects on lung and stomach cancer.

## 2.2 Ecologic Studies

A recent topic of research in the field of medical geography has been ecologic analyses of cancer. Ecologic analyses are a form of epidemiological research that use groups as the unit of study; for example counties or census districts. An epidemiological study is a statistical means of comparing the frequency of a particular effect in one group of people with that of another group or with the population as a whole. In the case of this study, it is an epidemiological study of the effects of socio-economic and environmental effects on people in Ontario counties.

According to the International Labour Office (1977), it should be possible in such a study to measure the level of exposure and occurrences of the effects so as to establish a dose-response relationship.

The use of the ecologic methodology has largely and historically been in the sociologic sciences, particularly the use of ecologic correlation analyses (Robinson, 1950). The fundamental research task of the disease ecologist is to investigate causes of disease by searching for associations between aspects of the social and physical environment and the disease under study (in this case cancer). If there is an association, the relationship must be studied further to see if the explanatory variable causes the dependent variable (Jones and Moon, 1987).

The ecologic approach has both advantages and disadvantages. One of the major positive aspects is that at low cost, large populations can be studied leading to the formation of hypotheses (Reynolds, 1989). There are many disadvantages to the ecologic approach that may lead researchers to dismiss results found. Many concerns are due to the quality and validity of the methodology. The collection of the data must be completed in an unbiased manner.

The major limitation of ecologic analysis for testing etiologic hypotheses is the potential for substantial bias estimating effects (Morgenstern, 1982). In

other words, can the results found for groups validly explain the situation at the individual level? This notion first demonstrated by Robinson in 1950 is known as the ecological fallacy which is made up of two parts. The first is the aggregate bias which is the aggregation or grouping of individual data and the second is the specification bias which comes from which characteristics of the group itself act as confounders ie. other risk factors are not homogeneous across different groups (Morgenstern 1982, Reynolds,1989).

Another problem with ecologic analyses is that of multi-collinearity of the independent variables. This occurs when the independent variables are highly correlated with each other especially socio-demographic and environmental variables (Leigh, 1988). The predictor variables are more highly correlated with each other than they are at the individual level. The increased correlations between the independent variables, make it difficult to determine their individual effects on stomach and lung cancer. It gives most problems to geographically defined units of analysis that are large and/or few in number (Morgenstern, 1982).

Since this multi-collinearity may seriously limit our ability to test an etiologic hypothesis with ecologic data, we must use methods to help minimize these problems. Such methods are a) using ecologic regression not

correlation to estimate the magnitude of the desired association, b) make the groups as homogeneous as possible (the independent variables) and c) assess how the groups were formed and analyze the data accordingly (Morgenstern, 1982).

Based on the limitations of the ecologic approach, it must be understood that the purpose of the study will only be to generate hypotheses of the environmental effects on cancer and not to establish causal relationships.

## 2.3 Disease Mapping

Thematic maps focus on displaying geographic occurrences and variations of certain phenomena. Cancer incidence rates can be mapped which makes visualizing geographic distributions much easier. A large volume of data can be visually comprehended almost immediately. Clustering of high rates in certain areas serve as "smoke signals" to environmental exposure that may be uncovered through further study (Blot et al., 1979). By revealing these clusters on disease maps, hypotheses can then be generated and tested further considering all potential reasons. Therefore, we see once again, as in the ecologic approach, that cancer maps serve to raise questions not answer them.

Disease maps tend to have some distortion due to the

use of artificial boundaries, although statistical data usually is available for these artificial boundaries (ie. county, census district).

The use of mapping cancer incidence rates in an ecologic analyses is to compare the mapped data to the geographic distribution of risk factors. The comparison may then be analyzed by a regression model to see how well the hypothetical risk factors predict the observed pattern of disease. Again, the results serve as hypothesis generating.

## 2.4 Environment and Health

### 2.4.1 Smoking

In Canada, smoking statistics are reported by age, sex, education and region. Cigarette smoking is clearly the major cause of lung cancer (Sebastini, 1983, U.S. Surgeon General, 1985) and in 1981 almost one third of Ontario residents over the age of 15 smoked. In 1986, 26% of the population in Ontario were smokers. From 1966-1986, the trend in smoking has decreased for males but there has been little change for females. the trend in the smoking behaviour of women has important implications for the future pattern of lung cancer mortality. Lung cancer mortality in females has been on the rise and is expected to surpass breast cancer as the leading cause of cancer deaths among

women (Millar, 1988).

The prevalence of regular smoking has been closely associated with the level of education. It was found that adults with secondary school level of education or less were twice as likely to be smokers than those with a university degree or 31% versus 15%. The importance of education is particularly evident among women. Those with lower levels of education were three times as likely to smoke than those with a university degree.

The importance of smoking within occupational groups is very important for epidemiological programs. Smoking may interact with hazards in the workplace in a "synergistic" manner to create more disease than the sum of the exposures separately (Millar, 1988). It is important to note that as of January 1989, smoking was completely banned in all federal public service settings as well as in many other workplaces since. The majority of Canadians favour more restrictive policies regarding smoking in public setting. The attitudes of Canadians toward smoking in the workplace may result in changes in the social environment that may lead to an acceleration in the decline of smoking in the future.

Regional variation in the smoking behaviour in Canada does exist but these differences are becoming smaller. We know that within each province, smoking rates are higher in males but in Ontario the amount of males that

smoke are lower than the national average. The amount of women in Ontario that smoke is on the rise. Unfortunately, there are no published data as of yet of the geographic distribution of smoking in Ontario.

## 2.5 Spatial Variations in Cancer

### 2.5.1 Cancer (Causes and Avoidable Risks)

Cancer is classified as a chronic degenerative disease (malignant neoplasm). It has been found that in economically advanced countries, chronic degenerative diseases are major killers especially over age 60. The importance of environmental factors on cancer occurrences is increasing. In the last 30 years, there has been a worldwide increase in lung cancer. The development of cancer depends on a multitude of variables including both characteristics of the individual (e.g., age, sex, genetic predisposition) and the environment in which one lives (e.g., dietary habits, lifestyle) (Reynolds, 1989).There is reasonable certainty that cancer is partially caused by the environment (Sebastini, 1983). It is suggested by many that possibly 90% of human cancers are determined by environmental factors (Doll and Peto, 1981). Sebastini (1983) reports that cancer deaths can be attributed to

various factors in the United States. Tobacco smoke is the most important cancerogenic factor at 30%. Diet accounts for 10-70% of cancer, occupational factors for 4%, pollution for 2% and industrial products 1%. Even though the author finds the attributable risk for occupation to be low at only 4%, it is the most important in social significance and especially has the potential for prevention.

### 2.5.2 Occupational Cancer

The fundamental task for occupational cancer epidemiology is to reveal causal connections between occupational exposures and various forms of cancer (Prevention of Occupational Cancer, 1981). There are problems and limitations in doing this. The long latency period between exposure and manifestation of the disease is the most difficult limitation (Occupational Cancer, 1977). There are also many uncontrolled factors that can contribute to making an epidemiological study on occupational cancer less informative. Hopefully, a dose-response relationship can be found to determine a causal relationship. The cancer incidence in the population of industrial societies is increasing.

According to the International Symposium on Prevention of Occupation Cancer, using an epidemiological

approach aids in the search for factors in the causation of cancer. Observations of the incidence of various cancers in occupational groups may provide clues for the detection of carcinogenic agents. Specific hypothesis can be tested by the observation of workers exposed to suspected carcinogens. Also, the occupational characteristics of affected individuals can be studied and compared to the population in general.

### 2.5.3 Environmental Effects on Cancer

Cancer was quite rare one hundred years ago, and primary lung cancer was practically unknown. It is now one of the most frequent causes of death and heads the list of mortality from cancer. Doll and Peto (1981) have considerable evidence to support links between exposure to carcinogens and cancer. However, the effects of the outside environment are more difficult to pinpoint due to such factors as the intensity of exposure that one experiences as well as the duration of exposure. It is much more complex to isolate factors of the environmental and their effects on cancer as compared to the occupational setting. It is also difficult to explain the effects of the environmental factors on cancer incidence as there may be many factors working simultaneously and it is very difficult to isolate them.

Epidemiological studies on death rates from lung

cancer suggest a close relationship between the incidence of cancer and steel making industrial centres (Cecilioni, 1976). Cecilioni also concluded that in cities, the death rate from lung cancer is about twice as high as in rural areas. Low concentrations of pollutants over extended periods of time can contribute substantially to a persons' total exposure to pollutants. Shannon et. al (1989) concluded in a more recent and thorough study on the relationship between air pollution and lung cancer in Hamilton that there has probably been some effect of air pollution on lung cancer mortality but the effect is lower than previously reported. Ethnicity including British, French, Dutch, German, Italian and Native consistently explained a substantial proportion of the variation of cancer mortality rates in Ontario.(Reynolds, 1988).

## 2.6 Conclusion

The continual increase in cancer rates and evidence of environmental effects on cancer mortality creates the need for serious concern from health care professionals. There are many areas that can benefit from an increasing body of knowledge on potential avoidable risks that can be undertaken in order to prevent the disease. The cause and cure of cancer have been researched extensively but no cure has been found. This paper will examine the possible

effects that the environment may play in the combination of causal agents of stomach and lung cancer. Incidence rates will be used as they provide more confidence of links with the environment (compared to mortality rates) because of the shorter time between exposure and diagnosis of incidence. No causal relationships can be made but potential links may be uncovered that could possibly result in future studies.

# CHAPTER 3 : RESEARCH METHODOLOGY

## 3.1 Research Hypothesis

This study is an ecologic analysis of cancer in Ontario counties. The purpose of the research is to examine the relationship between two of the most common cancer types (stomach and lung) in Ontario counties from 1976-1985 and environmental and socioeconomic related variables collected from the 1981 Census of Canada. Pinpointing potential causes of cancer from spatial distributions can help establish avoidable risks.

## 3.2 Data Sources

There are two data sources used for this research. First of all, the rates used for the cancer sites were standardized incidence rates for lung and stomach cancer. The rates were standardized by sex and age. The source of these rates was the Ontario Cancer Registry. The data was obtained from Dr. Stephen Walter from the McMaster University Medical Centre. The incidence rates (4) will serve as the dependent variables for the data analysis and will be mapped to determine possible clusters.

There are 17 ecologic variables that were obtained to serve as the independent variables for the analysis. The data for these variables were obtained from Master's Theses

done by D. Reynolds and C. Hampson. In height of previous findings, the 17 ecologic variables that were chosen included characteristics of socioeconomic status (education level, income), ethnicity ( ie. % french, % british etc.) and environmental variables (urban, population density, type of employment). They were drawn from the 1981 Census of Canada for each of the 47 census districts of Ontario. A list of the ecologic variables can be found on Table 2.1

For the disease mapping, the standardized incidence rates were used for each of the Ontario Counties and were entered into an Atlas Graphics Package.

## 3.3 Methods of Analysis

The standardized incidence rates were used along with the "Atlas Graphics" package to create a disease map of Cancer in Ontario counties. Separate maps were generated for males and females for each cancer site. These cancer incidence maps were created to reveal the existing distribution of stomach and lung cancer for men and women in each of the Ontario counties.

The statistical analysis was based on the SAS software package. Analysis were performed to generate descriptive statistics, correlation and multiple regression analyses. The descriptive statistics found for each set of variables were the maximum value, the minimum value, the standard deviation and the mean.

Table 2.1 : Ecologic Variables

1. urban (5000) %
2. manufacturing (1000 urban )#
3. population density
4. educational ( < grade 9) %
5. incidence of low income %
6. primary employment %
7. tertiary employment %
8. urban (census) %
9. median income $
10. secondary employment %
11. % british
12. % french
13. % dutch
14. % italian
15. % native
16. % german
17. manufacturing (1000) #

The correlation analysis tested the bivariate relationships among the 21 variables (cancer rates and 17 ecologic variables). These results were analyzed in order to screen the variables for use in the regression analysis. There were two purposes for performing the correlation analysis. The first purpose was to detect multicollinearity among ecologic variables. The second purpose was to identify strong predictors of disease rates. For example, it was established that the relationship between median income and low education was not very strong and therefore each could be used as an independent variable in the regression analysis.

The methodology used to examine the relationship between cancer rates in counties in Ontario and the environmental variables was multiple regression analysis. The stepwise procedure was used with backward elimination. Education <grade9 and median income (or low income) were forced to enter and remain in each equation to control for the effects of smoking. The dependent variables used were the standardized incidence rates for lung and stomach cancer, for men and women. The independent variables used were each of the 17 ecologic variables. The "best fit" equation for each of the cancer sites was determined after running approximately 27 multiple regression analyses. A list of all the equations after the backward elimination procedure was completed is included as the appendix.

# CHAPTER 4 : DATA ANALYSIS AND RESULTS

## 4.1 Introduction

Disease mapping of geographic distributions of cancer allow for a large volume of data to be visually comprehended almost immediately. The use of these disease maps is extensive.

## 4.2 Mapping of Dependent Variables (Cancer Sites)

Figures 1-4 illustrate the geographic distribution of cancer in Ontario by census district for each sex and site under study.

### 4.2.1 Lung Cancer, Female - Figure 1

The geographic distribution of female lung cancer incidence in Ontario had the highest rates in the northern part of the province. However, it must be kept in mind that the northern part of the province has larger counties in relation to area. Although the areas are larger, the populations are not and therefore the maps may give a misleading impression of the importance of the incidence rates of these areas. If the land area was proportional to population the picture would appear quite differently. The

counties with the highest rates (1.123 to 2.164) were Manitoulin, Sudbury, Rainy River and Thunder Bay. The surrounding northern provinces fall into the second highest category (1.006 to 1.123).

In Southern and Eastern Ontario, the largest rates are found in Toronto-York, Hamilton-Wentworth, Hastings and Frontenac. The lowest rates in the province seemed to be located in the region leading to, and forming, the Bruce Peninsula.

### 4.2.2 Lung Cancer, Male - Figure 2

The pattern for male lung cancer in Ontario was somewhat similar to that of females. The highest rates are located in northern Ontario, especially northwestern Ontario. The highest rates (1.097 to 2.241) are found in the northern counties of Manitoulin, Sudbury, Thunder Bay, Algoma,and Rainy River. The rate for males in the northeastern part of the province is significantly lower than that of females. In Southern Ontario, the highest incidence rates were found in Toronto-York and Wellington counties. There were pockets of fairly high rates in southwestern Ontario. In the eastern part of Ontario, the distribution of cancer was highest in Hastings and the surrounding counties were also fairly high.

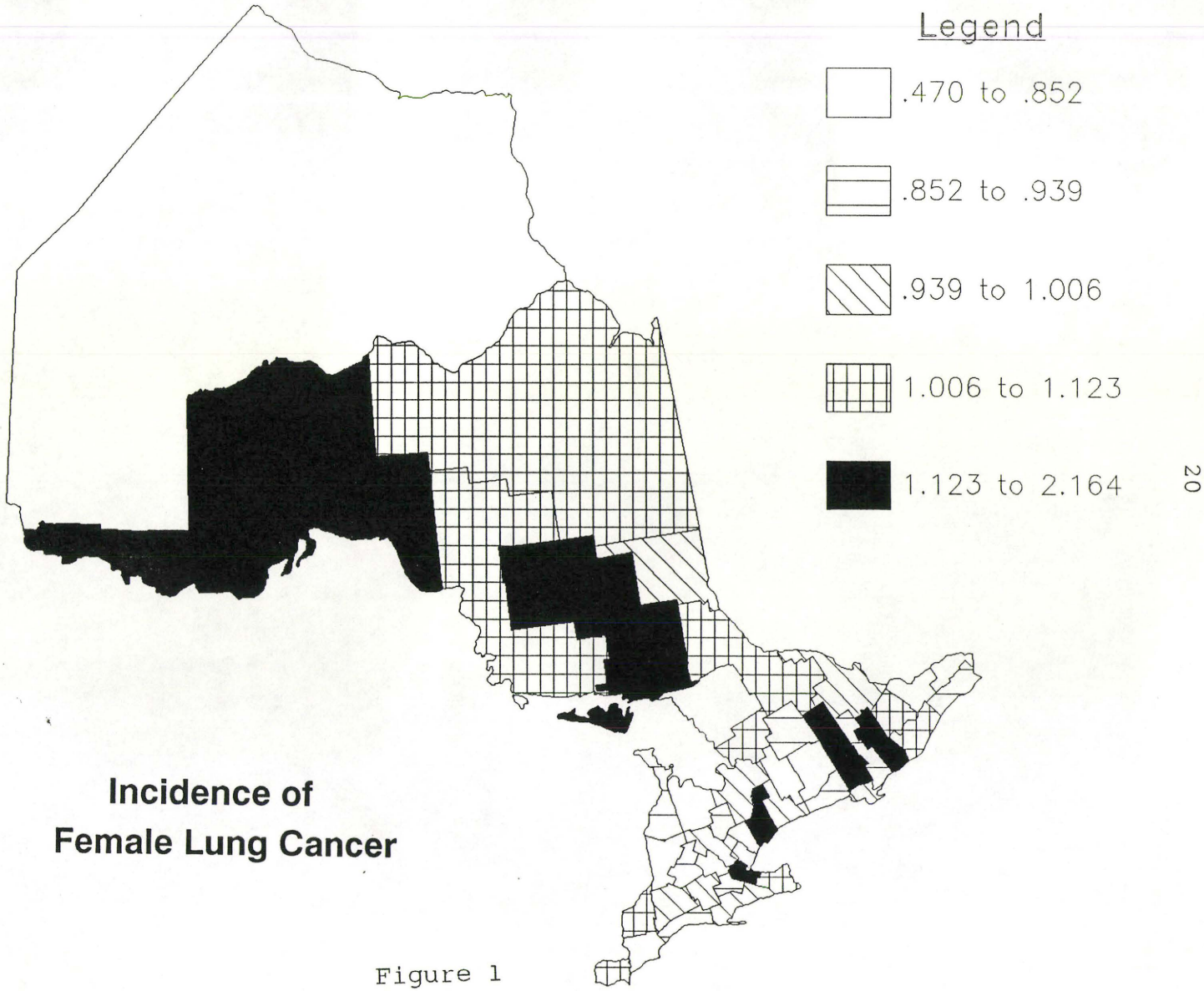

Legend
.470 to .852
.852 to .939
.939 to 1.006
1.006 to 1.123
1.123 to 2.164
Incidence of
Female Lung Cancer

Figure 1

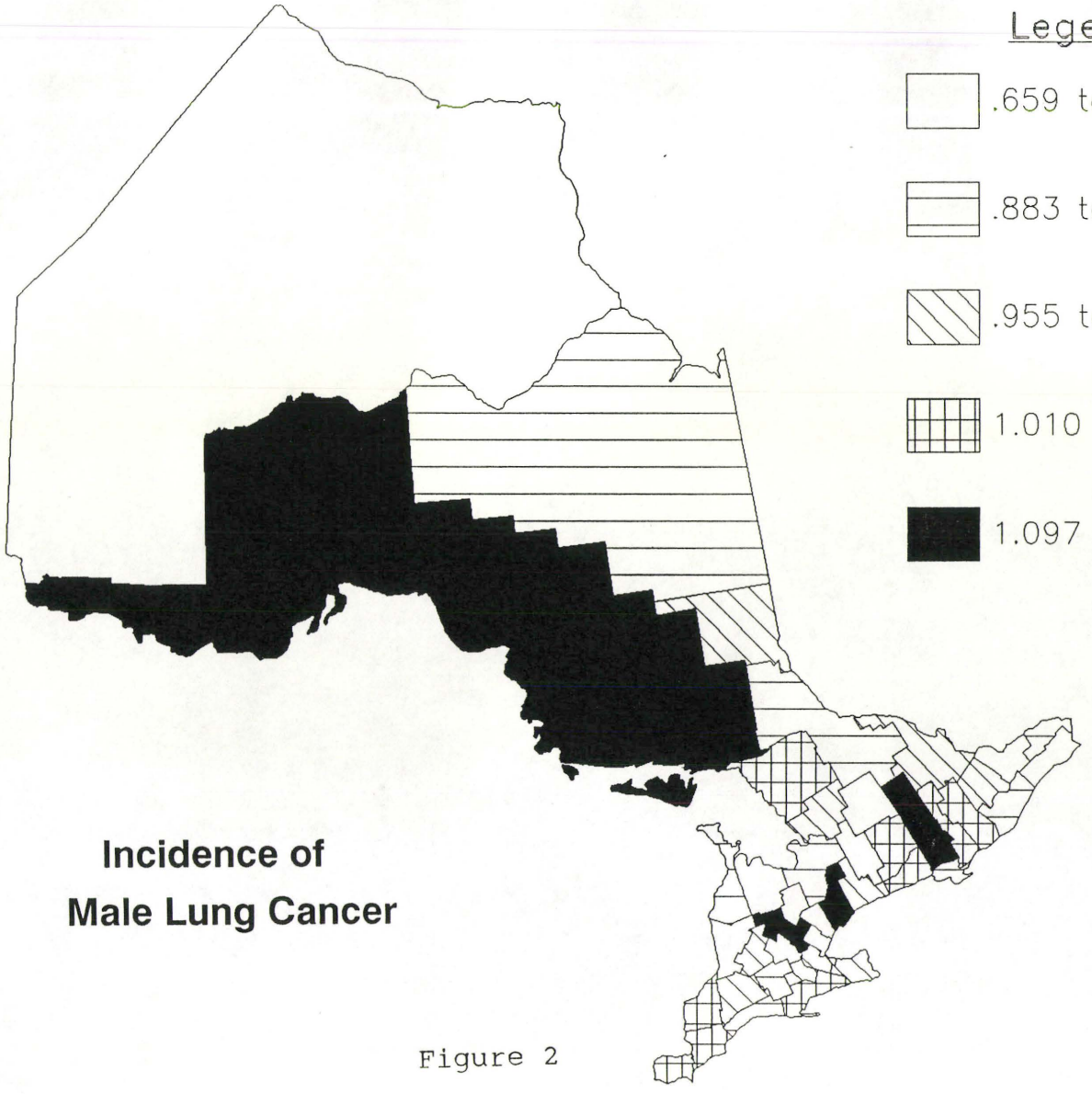

Legend
.659 to .883
.883 to .955
.955 to 1.010
1.010 to 1.097
1.097 to 2.241
Incidence of
Male Lung Cancer

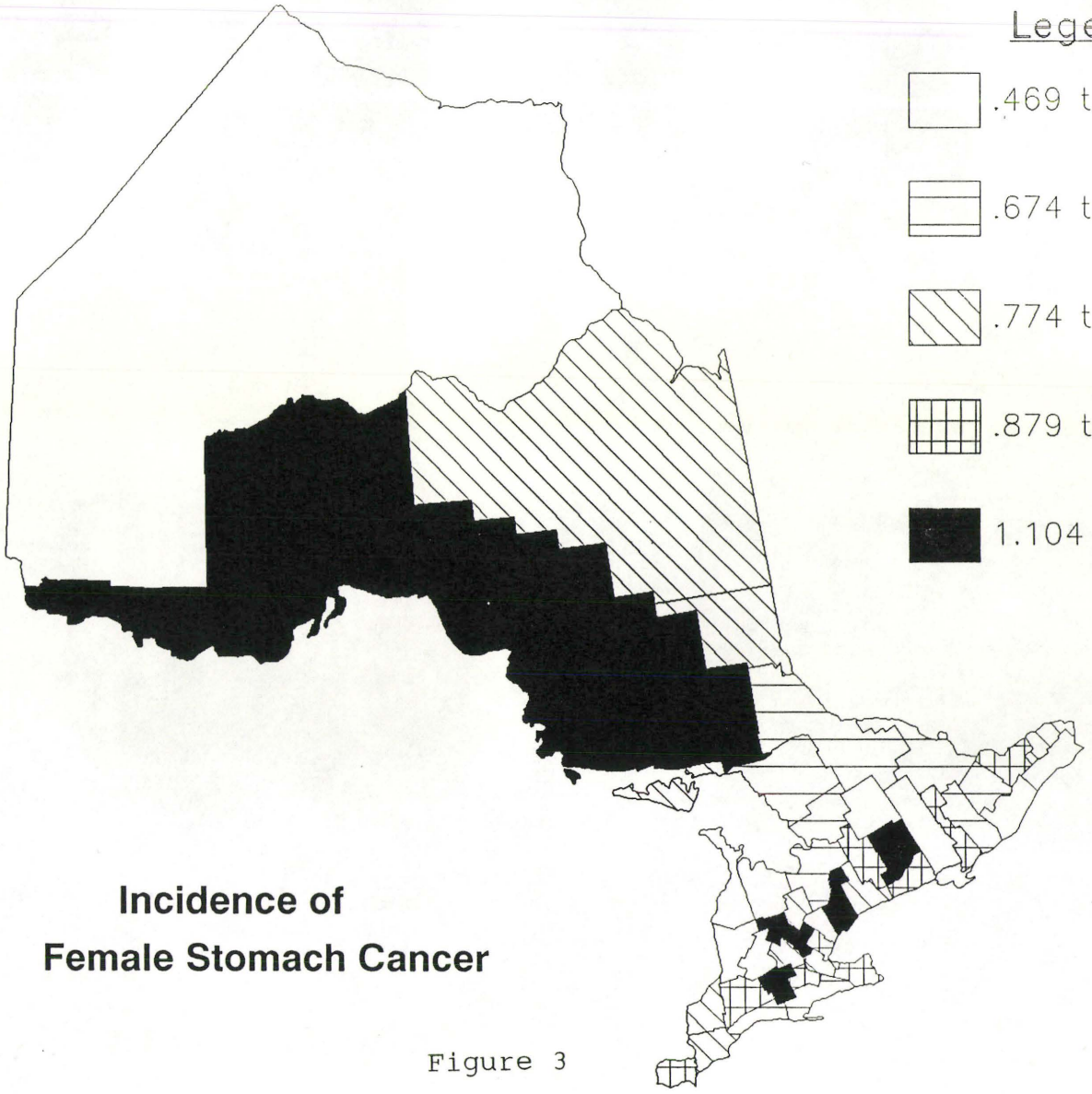
Figure 2

### 4.2.3 Stomach Cancer, Female - Figure 3

The pattern was most striking in northwestern Ontario for female stomach cancer. The counties in this area within the highest rate were Rainy River, Thunder Bay, Sudbury and Algoma. The rest of the province appeared to have consistently lower incidence rates except for a few exceptions in the southern part of the province. There were four counties that were within the highest rate of incidence. They were, Toronto-York, Peterborough, Wellington and Oxford. It is interesting to note that Hastings had a low incidence value with respect to female stomach cancer unlike the previous two maps studied.

### 4.2.4 Stomach Cancer, Male - Figure 4

The rates for male stomach cancer seem to be the highest in urban areas. Included in these urban counties are Toronto-York, Hamilton-Wentworth and Sudbury. The highest rates appear to be located once again in Northern Ontario but are not as strong as in the previous sites. The highest range of incidence rates are only from 1.110 to 2.391. Also, there are fewer counties in northern Ontario within this highest group. There are comparatively just as many in the second highest range at .968 to 1.110.




**Incidence of**
**Female Stomach Cancer**

Figure 3

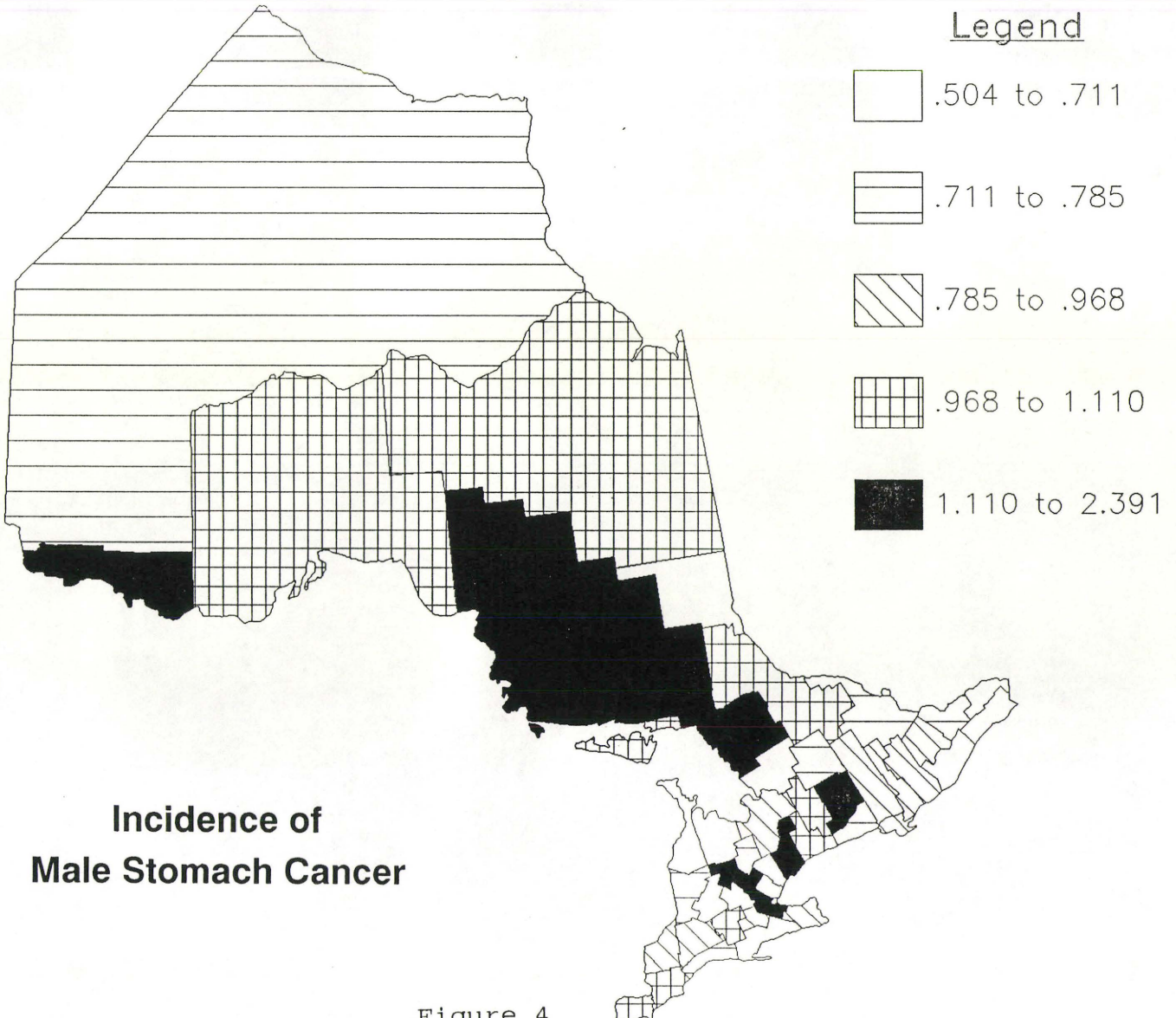

Legend
.504 to .711
.711 to .785
.785 to .968
.968 to 1.110
1.110 to 2.391
Incidence of
Male Stomach Cancer

Figure 4

## 4.3 Descriptive Statistics

Descriptive statistics were completed for all 21 variables. Included were the mean, standard deviation, minimum and maximum. The descriptive statistics were analyzed to see whether there were any discrepancies in the data. The average values for the lung cancer (men and women) and the values for the stomach cancer(men and women) were similar. The maximum value for urban %(5000) was 100% whereas the average in Ontario was only 56%. There was a very high value for population density at 1001 for Toronto-York but the value for the province was only 81. This reveals an area that is very densely populated. The rest of the values show no strong variations except for the ethnicity variables. It is interesting to note that the native population has a maximum value of 30.6 but an average value of 2.36. This means that the native population is very high in a certain area (e.g., Manitoulin) but very scarcely located elsewhere in the province. All of these variations need to be kept in mind for they may cause unrealistic high or low values in future analyses that may not be truly representative.

## 4.3 Correlation Analysis

The bivariate correlation analysis among the 21 variables shows the degree of linear association between

each pair of variables. The detection of multicollinearity can be determined by a high correlation value for two variables. The second purpose of the correlation analysis is to identify strong predictors of disease rates. The results of the correlation analyses are presented in Table 4.1. Pearson correlation coefficients were obtained for each pair of variables but not all are shown in the tables. The table includes the correlations between cancer outcomes and selected ecologic variables. This table provides the means to further assess results obtained in the multivariate analysis.

### 4.3.1 Interpretation of Correlation Analysis

There are two significant relationships between male lung cancer and the ecologic factors. They include % urban at .30 ($p<.05$) and population density at .39 ($p<.01$). The remaining correlations are not strong relationships.

There are similar findings for female lung cancer and the ecologic factors. The variables % urban and population density are the only two significant relationships at .31 ($p<.05$) and .41 ($p<.01$) respectively. The remaining ecologic factors do not have a strong relationship with female lung cancer.

Table 4.1

Correlations Between Cancer Outcomes and Ecologic Factors

| ECOLOGIC FACTORS | CANCER SITES | | | |
|---|---|---|---|---|
| | Lung Male | Lung Female | Stomach Male | Stomach Female |
| Median Income | .22 | .15 | .25 | .31[1] |
| Low Income | .10 | .18 | .09 | .01 |
| Education | .01 | .02 | .01 | -.08 |
| Urban % (census) | .30[1] | .31[1] | .33[1] | .40[2] |
| Population Density | .39[2] | .41[2] | .49[3] | .40[2] |
| Secondary Employment | -.14 | -.17 | -.18 | -.17 |
| Primary Employment | -.21 | -.25 | -.29[1] | -.25 |
| Urban (5000)% | .16 | .22 | .27 | .22 |
| Manufacturing # (1,000) | -.13 | -.14 | -.07 | -.11 |
| Manufacturing # (1,000 urban) | -.27 | -.24 | -.20 | -.33[1] |
| British % | -.19 | -.13 | -.11 | -.19 |
| Italian % | .27 | .19 | .36[2] | .28[1] |
| Native % | .02 | .05 | .05 | -.06 |

1 $p < 0.05$
2 $p < 0.01$
3 $p < 0.001$

Male stomach cancer has four significant relationships with ecologic factors. Once again, % urban with the correlation value of .33 ($p<.05$) and population density at .49 ($p<.001$) are strongly related. Primary employment has a negative significant relationship at -.29 ($p<.05$). This means that as the amount of primary employment increases, the male stomach rate decreases. The final significant relationship is between male stomach cancer and % italian with a value of .36 $p<.01$). This is consistent with previous findings by a thesis done by D. Reynolds. As the % italian increases, the rate of male stomach cancer increases.

Female stomach cancer has five significant relationships. Once again the % urban and the population density variables are among the highest values at .40 ($p<.01$) for both. Median income is positively related with a value of .31 ($p<.05$). The number in manufacturing (1,000 urban) has a negative significant relationship at -.33 ($p<.05$) indicating that as the number per 1,000 in manufacturing increases, the female stomach cancer rate decreases. Finally, the % italian is positively related to female stomach cancer with a correlation of .28 ($p<.05$). As the % italian increases, the rate of female stomach cancer increases.

The correlation coefficients can be summarized according to ethnicity, environmental variables and socio-

economic status. The ethnicity variables (% british, italian and native) have two significant relationships. Both are stomach cancer for male and female with % italian. It can be assumed according to the data, ethnicity does not have a strong bearing on cancer rates, except for italian.

The environmental factors (urban %, population density, type of employment) include ten significant relationships. Attention must be brought to the strong relationships between all cancer sites and % urban and population density. For both stomach and lung cancer for males and females, as the % urban and population density increase, the cancer rates increase. The two other strong relationships are related to the type of employment and are both negative.

The socio-economic variables (income and education factors) contain only one significant relationship. It is the female stomach cancer and median income. Therefore, it may be said that socio-economic status and cancer are not strongly related.

Previous studies have shown that smoking is strongly correlated with socio-economic status (income and education) at the individual level. The data shows no effect for education and a weak effect for income. This is aggregate data and therefore, inferences cannot be make about individuals. It is still important in this study to include income and education in the regression analysis to control

for the effects that smoking will have.

## 4.5 Multivariate Analysis

### 4.5.1 Description of Regression Analysis

Multiple regression analyses were run for many different combinations of independent variables. For each analysis, the low income and education variables were forced to remain in the equation to control for the effects of smoking. The regression analysis that was used was a multiple regression with the stepwise procedure using backward elimination. As the independent variables became insignificant in the regression, they were removed (except for the income and education variables that were kept whether significant or not). Table 4.2 shows the final "best fit" regression equations.

### 4.5.2 Interpretation of the Multiple Regression

For male lung cancer, the "best fit" equation included 6 variables, of which 4 (education, urban % census, population density and manufacturing # per 1,000) were significant at the 0.05 level. The regression coefficients indicate that male lung cancer rates are positively related to median income, education, urban % (census), population

density and % British and negatively related to manufacturing # (1,000). Overall, the equation accounts for 39 % of the total variation in male lung cancer rates.

For Male Lung Cancer, the equation included median income and education to control for the effects of smoking. It was found that urban % was significant at the .05 level. As the census urban population increases by 1%, male lung cancer increases by .0076. Population density was also found to be significant. As the population density increases by 1, the male lung cancer rate increases by .008. As the # per 1,000 manufacturing increases by 1,000, male lung cancer decreases by .305. British % was also included in the "best fit" equation but was not significant at the $p<.05$ level. The intercept was not significant at the $p<.05$ level.

For female lung cancer, the "best fit" equation included 6 variables of which 5 (education, urban % census, population density, manufacturing # per 1,000 and % British) were significant at the 0.05 level. The regression coefficients indicate that female lung cancer rates are positively related to education, urban % (census), population density and % British and negatively related to median income and manufacturing # (1,000). Overall, the equation accounts for 42 % of the total variation in female lung cancer rates.

Table 4.2

Summary of Regression Analysis

| CANCER SITE | "BEST FIT" EQUATION |
|---|---|
| Male Lung | $Y = -.73 + .00001$ median income $+ .045^{1}$ education $+ .008^{1}$ urban %(census) $+ .001$ population density $- .305^{2}$ manufacturing #(1,000) $+ .009$ british% $+ E$<br>$R^{2} = .3899$ overall $F=4.15$ ($p < .01$) |
| Female Lung | $Y = -.41 - .000002$ median income $+ .04^{1}$ education $+ .008^{1}$ urban %(census) $.001^{2}$ population density $- .32^{2}$ manufacturing #(1,000) $+ .009^{1}$ british% $+ E$<br>$R^{2} = .4211$ overall $F=4.73$ ($p < .001$) |
| Male Stomach | $Y = -.2.44^{1} + .00004$ median income $+ .074^{3}$ education $+ .010^{2}$ urban %(census) $+ .001^{2}$ population density $- .36^{3}$ manufacturing #(1,000) $+ .016^{2}$ british% $+ E$<br>$R^{2} = .5278$ overall $F=7.27$ ($p < .001$) |
| Female Stomach | $Y = .522 + .000023$ median income $+ .015^{2}$ education $+ .0015$ population density $- .354^{1}$ manufacturing #(1,000) + Error<br>$R^{2} = .2867$ overall $F=4.12$ ($p < .01$) |

1 $p < 0.05$
2 $p < 0.01$
3 $p < 0.001$

For female lung cancer, median income and education were also used to control for the effects of smoking. As urban %(census) increases by 1%, female lung cancer increases by .008. As the population density increases by one, the female lung cancer rate increases by .001.

As manufacturing #(1,000) increases by 1,000, the lung cancer rate decreases by .32. In this equation, British % was significant at the $p<.05$ level. As the percentage British increases by 1%, female lung cancer increases by .009. In this equation, the intercept was not found to be significant at the $p<.05$ level therefore, the independent variables are relevant in explaining the rate of incidence of female lung cancer.

For male stomach cancer, the "best fit" equation included 6 variables of which 5 (education, urban % census, population density, manufacturing # per 1,000 and %British) were significant at 0.05 level. The regression coefficients indicate that male stomach cancer rates are positively related to median income, education, urban % (census), population density and % British and negatively related to manufacturing # (1,000). Overall, the equation accounts for 53% of the total variation in male stomach cancer rates.

Median income and education were again used to control for the effects of smoking on male stomach cancer and both were significant at the 0.05 level. For the percentage urban, as it increases by 1%, the male stomach

cancer increases by .010. As population density increases by one, the stomach cancer rate increases by .001. As the # manufacturing increases by 1,000, the male stomach cancer rate decreases by .36. As the % British increases by 1%, the cancer rate increases by .016. The intercept was found to be significant at the $p<.05$ level and therefore it can be said that there are other independent factors relevant besides those used in this study to explain the incidence of male stomach cancer in Ontario.

For female stomach cancer, the "best fit" equation included 4 variables, of which 2 (population density and manufacturing # per 1,000) were significant at the 0.05 level. The regression coefficients indicate that female stomach cancer rates are positively related to median income, education, and population density and negatively related to manufacturing # (1,000). Overall, the equation accounts for 29 % of the total variation in female stomach cancer rates.

For female stomach cancer, median income and education were also used to control for smoking. After the regression, neither variable was significant at the $p<.05$ level. As population density increases by 1, the female stomach cancer rate increases by .0015. As the # manufacturing increases by 1,000, the female stomach cancer rate decreases by .354. These were the only variables that remained to be significant after the stepwise procedure.

# CHAPTER 5: CONCLUSIONS

## 5.1 Summary of Findings

In summary, the analysis and results lead to various conclusions. The purpose of examining the effects of certain environmental, ethnic and socio-economic variables on cancer incidence in Ontario has been completed by mapping, correlation and regression analyses. All the "best fit" regression equations used median income and education less than grade 9 to control for the effects of smoking.

Male lung cancer incidence had the highest rates in northwestern Ontario. High rates were also found in Toronto-York and Wellington counties. Both the correlation and regression analyses revealed the positive relationships with population density and % urban.

Female lung cancer distributions were similar to male lung cancer. The northwestern part of the province contained the highest incidence rates with selected highs in southern Ontario. Counties such as Hamilton-Wentworth and Toronto-York and Sudbury were included in the highest quintiles. This strengthens the correlation and regression results which have % urban and population density significant in both analyses. Other highly urban areas also had high incidence rates such as Windsor and Ottawa. This may stimulate further study into the effects of the

environment on cancer. The other significant variable was % british which does not coincide with a map of the British distributions in Ontario (Reynolds, 1989).

Male stomach cancer incidence was highest in northern Ontario (but not as high as lung cancers) but the most striking distribution coincided with urban areas. The areas include Hamilton-Wentworth, Toronto-York, and Sudbury once again. According to the results of the regression analysis, this relationship cannot be linked to manufacturing in these areas because the manufacturing rate is negatively significant. As the manufacturing % increases by 1,000, the stomach cancer in males decreases by .016. Once again, the % urban and population density variables were significantly related to the cancer incidence inferring a environmental effect. The final relationship in the correlation analysis was the % italian but was not significant in the regression analysis.

Female stomach cancer incidence also shows the highest distributions in northwestern Ontario with a smaller amount of occurrence in urban areas. Both the regression and correlation analyses show a strong relationship between population density but % urban was only significant in the correlation analysis. Manufacturing % has a strong negative relationship in correlation and regression results. The % italian is significant for female stomach cancer ( as seen in the correlation analysis only).

Therefore, there is a tendency toward the strongest relationships being between the cancer incidence and selected environmental variables. Manufacturing has a negative significant rate for each site for both sexes. The percentage italian shows a relationship with stomach cancer in males and females.

## 5.2 Consistency of the Results

The results of this study are fairly consistent with the findings of past studies. According to the mapping, the analysis showed that northern Ontario had the highest rates. Reynold's (1989) disease mapping using mortality rates also showed the highest rates for lung and stomach cancer for males and females to be in the northern part of the province. She also found the results for the male stomach cancer to be dominant in urban areas. This is the case in this study adding the relevance of urban living for both cancer sites. The results of the analysis also showed % italian to be important in male and female stomach cancer which is consistent with previous findings.

The cancer clusters are evident in Ontario for male and female lung and stomach cancer. This is consistent with Walter's study where significant clusters by cancer site were found for male and female lung and stomach cancer. The results also showed that environmental factors played an

important role in each of the cases. This result is consistent with Doll and Peto (1981) in which they stated that possibly 90% of human cancers are determined by environmental factors. Although this is the case, the county of Hamilton-Wentworth was only amongst the highest in lung cancer rates for females and not males. This is fairly consistent with Shannon (1976) in which he concludes that air pollution effects in Hamilton are evident but much lower than previously reported.

## 5.3 Research Directions

Ecologic studies of cancer in Ontario may be completed more thoroughly in the future due to the release of smoking data for the province that will be at the individual level. This will allow for the effects of smoking to be more detailed and accurate. The results of this study showed percentage urban and population density to strongly related to the incidence of cancer in Ontario. This could lead to future studies of urban areas in which factors could be isolated to determine their contributions. This would be difficult to conduct due to the potential effects factors have when acting together. This study also made a relationship to the rates of stomach cancer and italians. This could be studied further testing what exactly they are consuming that makes their levels higher.

Appendix I

## DATA FOR VARIABLES

This list include the data for all 47 county divisions used for both the standardized incidence rates and the 17 ecologic variables.

| OBS | COUNTY | MALE LUNG | FEM LUNG | MALE STOM | FEM STOM | URB01 | URB02 | POPDEN | EDUC | LOWINC |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ALGOMA | 1.546 | 1.114 | 1.561 | 1.347 | 74.44 | 74.52 | 2.6 | 19.43 | 10.2 |
| 2 | BRANT | 0.980 | 0.939 | 0.833 | 0.962 | 78.33 | 79.49 | 95.6 | 19.62 | 11.5 |
| 3 | BRUCE | 0.886 | 0.873 | 0.706 | 0.655 | 19.84 | 42.42 | 14.8 | 21.35 | 11.2 |
| 4 | COCHRANE | 0.946 | 1.095 | 1.110 | 0.869 | 72.27 | 69.01 | 0.7 | 26.35 | 11.1 |
| 5 | DUFFERIN | 0.741 | 0.902 | 0.731 | 0.832 | 44.12 | 57.24 | 20.9 | 15.89 | 8.9 |
| 6 | DURHAM | 0.980 | 0.959 | 1.034 | 0.879 | 88.03 | 82.14 | 113.9 | 13.65 | 8.3 |
| 7 | ELGIN | 0.885 | 0.865 | 0.732 | 0.723 | 47.95 | 54.06 | 37.1 | 20.45 | 10.9 |
| 8 | ESSEX | 1.040 | 1.010 | 0.980 | 0.929 | 72.99 | 80.10 | 167.8 | 19.77 | 13.9 |
| 9 | FRONTENAC | 1.067 | 1.320 | 0.824 | 0.765 | 48.66 | 72.08 | 28.3 | 14.43 | 12.5 |
| 10 | GREY | 0.722 | 0.710 | 0.647 | 0.612 | 35.49 | 50.12 | 16.4 | 22.82 | 13.9 |
| 11 | HALD-NORFLK | 1.045 | 0.955 | 0.711 | 0.708 | 69.71 | 42.18 | 30.7 | 23.13 | 10.4 |
| 12 | HALIBURTON | 0.659 | 0.909 | 0.726 | 0.657 | . | . | 2.7 | 23.66 | 15.6 |
| 13 | HALTON | 0.883 | 0.990 | 0.785 | 1.054 | 100.00 | 91.48 | 264.8 | 9.33 | 6.3 |
| 14 | HAM-WENTWRTH | 1.010 | 1.154 | 1.117 | 0.872 | 91.68 | 91.22 | 369.7 | 20.16 | 14.0 |
| 15 | HASTINGS | 1.250 | 1.152 | 0.843 | 0.674 | 46.75 | 64.73 | 17.9 | 18.33 | 13.1 |
| 16 | HURON | 0.688 | 0.470 | 0.740 | 0.482 | 13.04 | 35.83 | 16.5 | 21.96 | 13.2 |
| 17 | KENORA | 0.718 | 0.600 | 0.713 | 0.469 | 27.70 | 44.81 | 0.1 | 26.13 | 7.6 |
| 18 | KENT | 1.055 | 0.852 | 0.986 | 0.810 | 49.02 | 64.48 | 42.9 | 21.61 | 12.6 |
| 19 | LAMBTON | 1.097 | 1.015 | 0.790 | 0.776 | 41.23 | 69.03 | 41.2 | 15.49 | 9.8 |

| OBS | MEDINC | MFRG07 | MFRG08 | PRIMEMP | SECONEMP | TERMPL | BRISH | FRENCH | DUTCH | GERMAN | ITALIAN | NATIVE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 23957 | 0.57 | 0.79 | 0.44 | 2.22 | 55.07 | 46.17 | 15.22 | 0.88 | 2.75 | 9.30 | 3.18 |
| 2 | 20907 | 2.03 | 2.55 | 2.29 | 7.46 | 54.40 | 62.19 | 2.65 | 3.24 | 3.78 | 2.95 | 4.59 |
| 3 | 19127 | 1.20 | 2.82 | 3.05 | 3.94 | 50.39 | 64.72 | 3.22 | 2.67 | 14.21 | 0.36 | 1.87 |
| 4 | 21990 | 0.57 | 0.79 | 0.99 | 1.90 | 53.65 | 24.69 | 49.25 | 0.53 | 1.43 | 2.49 | 6.08 |
| 5 | 22579 | 1.74 | 3.03 | 2.15 | 4.80 | 55.40 | 74.09 | 2.02 | 4.30 | 3.39 | 1.12 | 0.36 |
| 6 | 25885 | 1.29 | 1.57 | 0.70 | 5.11 | 60.55 | 67.43 | 3.48 | 3.44 | 3.13 | 2.17 | 0.32 |
| 7 | 19717 | 1.45 | 2.68 | 5.57 | 6.29 | 48.58 | 65.88 | 2.46 | 5.78 | 7.66 | 0.74 | 0.31 |
| 8 | 20934 | 1.70 | 2.12 | 1.34 | 4.59 | 58.02 | 41.65 | 15.36 | 1.13 | 4.83 | 7.30 | 0.24 |
| 9 | 20032 | 0.87 | 1.21 | 0.48 | 2.24 | 64.86 | 70.09 | 4.57 | 2.56 | 2.72 | 1.13 | 0.28 |
| 10 | 16642 | 1.73 | 3.46 | 1.92 | 3.91 | 33.16 | 73.63 | 1.61 | 2.58 | 9.58 | 0.23 | 0.21 |
| 11 | 20316 | 1.44 | 3.42 | 9.40 | 4.00 | 44.13 | 59.46 | 2.26 | 5.58 | 7.72 | 0.66 | 0.97 |
| 12 | 14044 | 1.05 | . | 0.66 | 2.73 | 57.60 | 79.02 | 2.46 | 1.48 | 3.94 | 0.67 | 0.18 |
| 13 | 29496 | 1.80 | 1.97 | 0.64 | 3.78 | 71.03 | 63.17 | 3.41 | 3.49 | 4.02 | 2.91 | 0.16 |
| 14 | 21766 | 1.40 | 1.53 | 0.57 | 4.99 | 58.32 | 53.47 | 3.38 | 2.75 | 3.74 | 9.45 | 0.68 |
| 15 | 18468 | 1.71 | 2.64 | 0.72 | 6.99 | 59.85 | 71.94 | 5.04 | 2.71 | 3.16 | 0.75 | 1.75 |
| 16 | 17362 | 1.78 | 4.98 | 4.14 | 3.96 | 47.46 | 69.56 | 2.52 | 7.58 | 8.42 | 0.14 | 0.14 |
| 17 | 21129 | 0.75 | 1.52 | 1.02 | 1.63 | 57.10 | 34.34 | 6.96 | 0.89 | 5.06 | 1.19 | 26.12 |
| 18 | 19488 | 1.68 | 2.61 | 2.86 | 5.28 | 52.52 | 59.74 | 9.77 | 6.22 | 2.59 | 1.28 | 0.62 |
| 19 | 23136 | 1.02 | 1.48 | 1.53 | 2.69 | 57.49 | 64.59 | 5.65 | 5.33 | 3.02 | 1.88 | 2.44 |

| OBS | COUNTY | MALELUNG | FEMLUNG | MALESTOM | FEMSTOM | URB01 | URB02 | POPDEN | EDUC | LOWINC |
|---|---|---|---|---|---|---|---|---|---|---|
| 20 | LANARK | 0.964 | 1.052 | 0.822 | 0.694 | 44.05 | 54.73 | 14.9 | 17.62 | 11.0 |
| 21 | LEEDS-GREN | 0.928 | 1.123 | 0.602 | 0.586 | 24.59 | 44.22 | 23.9 | 15.82 | 10.1 |
| 22 | LENNOX-ADD | 1.079 | 0.921 | 0.866 | 1.104 | . | 33.03 | 11.6 | 18.25 | 12.1 |
| 23 | MANITOULIN | 1.144 | 1.348 | 1.045 | 0.819 | . | 13.70 | 3.0 | 30.69 | 16.4 |
| 24 | MIDDLESEX | 0.996 | 0.943 | 0.893 | 0.925 | 82.67 | 86.56 | 94.7 | 13.96 | 11.6 |
| 25 | MUSKOKA | 0.983 | 1.037 | 0.653 | 0.763 | 75.74 | 42.51 | 9.5 | 18.82 | 12.0 |
| 26 | NIAGARA | 0.991 | 1.059 | 0.968 | 0.923 | 95.70 | 87.53 | 199.0 | 19.22 | 12.3 |
| 27 | NIPISSING | 0.942 | 1.090 | 1.044 | 0.724 | 71.41 | 74.43 | 4.5 | 20.99 | 14.0 |
| 28 | NORTHUMBERLND | 1.016 | 0.933 | 0.780 | 0.990 | 32.91 | 47.99 | 30.8 | 18.07 | 10.0 |
| 29 | OTTAWA-CARL | 0.825 | 0.745 | 0.765 | 0.934 | 89.77 | 89.72 | 198.3 | 10.56 | 10.9 |
| 30 | OXFORD | 0.891 | 1.006 | 1.031 | 1.108 | 53.06 | 59.95 | 42.3 | 21.48 | 9.9 |
| 31 | PARRY SOUND | 1.052 | 0.841 | 1.284 | 0.727 | 18.26 | 28.51 | 3.3 | 24.81 | 15.5 |
| 32 | PEEL | 0.804 | 0.657 | 0.671 | 0.519 | 100.00 | 88.91 | 400.4 | 11.48 | 7.5 |
| 33 | PERTH | 0.967 | 0.885 | 0.704 | 0.554 | 47.34 | 61.14 | 30.2 | 21.51 | 10.4 |
| 34 | PETERBOROUGH | 1.069 | 0.816 | 1.130 | 1.244 | 59.17 | 66.68 | 25.9 | 16.20 | 12.1 |
| 35 | PRESCOTT-RUSS | 0.935 | 0.800 | 0.504 | 0.776 | 18.72 | 41.77 | 26.4 | 26.64 | 12.6 |
| 36 | PRINCE EDWARD | 0.955 | 0.959 | 0.917 | 0.672 | . | 24.37 | 21.3 | 21.17 | 11.5 |
| 37 | RAINY RIVER | 1.318 | 1.165 | 1.153 | 1.220 | 39.06 | 62.75 | 1.4 | 22.23 | 9.4 |
| 38 | RENFREW | 0.983 | 0.986 | 0.779 | 0.687 | 44.29 | 57.40 | 11.4 | 22.52 | 12.7 |

| OBS | MEDINC | MFRG07 | MFRG08 | PRIMEMP | SECONEMP | TEREMPL | BRITISH | FRENCH | DUTCH | GERMAN | ITALIAN | NATIVE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 19466 | 1.79 | 3.28 | 1.14 | 6.82 | 59.97 | 80.96 | 4.81 | 1.08 | 1.80 | 0.35 | 0.22 |
| 21 | 19479 | 1.32 | 2.99 | 1.59 | 6.54 | 57.75 | 76.48 | 4.99 | 3.89 | 2.29 | 0.44 | 0.25 |
| 22 | 19297 | 0.70 | 2.11 | 1.41 | 4.64 | 57.24 | 74.51 | 3.75 | 3.58 | 3.27 | 0.33 | 0.37 |
| 23 | 13020 | 0.82 | 6.00 | 1.61 | 1.49 | 55.33 | 56.28 | 2.57 | 0.55 | 0.73 | 0.27 | 30.61 |
| 24 | 20794 | 1.35 | 1.55 | 1.13 | 4.42 | 88.90 | 63.97 | 2.71 | 4.23 | 4.14 | 2.35 | 1.05 |
| 25 | 16940 | 1.38 | 3.25 | 0.71 | 3.04 | 59.82 | 75.41 | 3.53 | 1.25 | 4.05 | 0.65 | 0.86 |
| 26 | 21460 | 1.41 | 1.62 | 1.29 | 3.80 | 58.10 | 50.16 | 6.48 | 4.18 | 6.63 | 8.18 | 0.37 |
| 27 | 19338 | 1.08 | 1.46 | 0.55 | 2.75 | 64.30 | 40.49 | 33.03 | 0.65 | 2.86 | 2.09 | 1.78 |
| 28 | 19454 | 1.63 | 3.40 | 2.11 | 6.39 | 53.09 | 78.33 | 2.68 | 3.80 | 2.18 | 0.57 | 0.51 |
| 29 | 24860 | 0.63 | 0.70 | 0.25 | 1.45 | 82.95 | 48.67 | 20.86 | 1.19 | 2.69 | 2.89 | 0.30 |
| 30 | 20204 | 2.11 | 3.51 | 4.44 | 5.80 | 50.75 | 65.02 | 2.21 | 7.31 | 7.21 | 1.04 | 0.24 |
| 31 | 14348 | 1.28 | 4.48 | 0.64 | 2.13 | 57.69 | 68.63 | 5.30 | 1.64 | 5.60 | 1.24 | 2.28 |
| 32 | 29189 | 2.25 | 2.53 | 0.30 | 6.06 | 67.12 | 53.24 | 3.05 | 1.85 | 3.10 | 7.83 | 0.38 |
| 33 | 19280 | 1.77 | 2.90 | 3.88 | 7.88 | 49.90 | 59.12 | 1.67 | 4.58 | 17.23 | 0.87 | 0.21 |
| 34 | 19804 | 1.22 | 1.83 | 0.86 | 4.17 | 62.91 | 78.04 | 2.83 | 2.06 | 2.38 | 0.98 | 1.05 |
| 35 | 20154 | 1.44 | 3.45 | 2.48 | 4.72 | 54.13 | 16.73 | 75.83 | 0.87 | 0.84 | 0.36 | 0.11 |
| 36 | 18195 | 1.61 | 6.67 | 3.82 | 3.77 | 56.60 | 77.89 | 2.09 | 4.15 | 2.76 | 0.70 | 0.56 |
| 37 | 20176 | 0.75 | 1.52 | 1.94 | 2.67 | 55.17 | 43.81 | 6.52 | 2.17 | 4.98 | 0.77 | 9.18 |
| 38 | 18319 | 1.35 | 2.35 | 1.12 | 4.75 | 59.74 | 51.42 | 10.93 | 1.49 | 12.72 | 0.23 | 0.78 |

| OBS | COUNTY | MALELUNG | FEMLUNG | MALESTOM | FEMSTOM | URB01 | URB02 | POPDEN | EDUC | LOWINC |
|---|---|---|---|---|---|---|---|---|---|---|
| 39 | SIMCOE | 0.945 | 0.953 | 0.836 | 0.774 | 44.10 | 61.82 | 46.5 | 17.26 | 10.2 |
| 40 | STORMONT ETC. | 0.852 | 0.880 | 0.616 | 0.502 | 31.33 | 56.95 | 30.6 | 22.95 | 14.8 |
| 41 | SUDBURY | 2.241 | 2.158 | 1.860 | 3.223 | 37.63 | 82.36 | 4.1 | 22.14 | 13.3 |
| 42 | THUNDER BAY | 1.204 | 1.161 | 1.072 | 1.447 | 73.04 | 82.97 | 1.4 | 19.73 | 8.9 |
| 43 | TIMISKAMING | 0.994 | 0.974 | 0.650 | 0.839 | 43.05 | 66.70 | 3.2 | 26.11 | 15.6 |
| 44 | VICTORIA | 0.825 | 0.762 | 0.972 | 0.891 | 28.42 | 35.36 | 15.6 | 20.31 | 10.6 |
| 45 | WATERLOO | 0.870 | 0.808 | 0.689 | 0.847 | 87.18 | 91.23 | 224.7 | 19.20 | 10.6 |
| 46 | WELLINGTON | 1.154 | 0.994 | 1.280 | 1.130 | 59.70 | 70.59 | 48.7 | 17.20 | 9.1 |
| 47 | YORK-TORONTO | 2.186 | 2.164 | 2.391 | 2.630 | 98.52 | 98.06 | 1001.6 | 17.49 | 12.5 |

| OBS | MEDINC | MFRG07 | MFRG08 | PRIMEMP | SECONEMP | TERMPL | BRISH | FRENCH | DUTCH | GERMAN | ITALIAN | NATIVE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 39 | 20261 | 1.35 | 2.18 | 1.22 | 6.32 | 60.12 | 68.99 | 7.00 | 3.00 | 3.07 | 1.68 | 0.78 |
| 40 | 17787 | 1.31 | 2.29 | 1.94 | 5.54 | 54.63 | 46.84 | 32.78 | 3.27 | 1.92 | 0.61 | 0.83 |
| 41 | 21066 | 0.56 | 0.68 | 0.51 | 1.62 | 58.27 | 32.75 | 35.25 | 0.67 | 2.26 | 4.68 | 1.81 |
| 42 | 24265 | 0.80 | 0.96 | 0.66 | 2.33 | 58.37 | 38.96 | 8.34 | 1.38 | 2.78 | 6.78 | 3.78 |
| 43 | 16291 | 1.19 | 1.78 | 2.48 | 1.61 | 55.20 | 46.58 | 30.34 | 0.76 | 2.70 | 0.87 | 1.19 |
| 44 | 17465 | 1.42 | 4.02 | 1.43 | 5.75 | 52.45 | 80.76 | 2.36 | 2.16 | 2.44 | 0.51 | 0.34 |
| 45 | 22022 | 2.44 | 2.92 | 0.56 | 8.88 | 59.08 | 45.15 | 3.11 | 1.92 | 21.26 | 1.10 | 0.24 |
| 46 | 21759 | 2.06 | 2.92 | 1.73 | 6.73 | 58.54 | 62.82 | 1.84 | 4.72 | 6.80 | 5.29 | 0.20 |
| 47 | 26605 | 2.38 | 2.43 | 0.16 | 5.53 | 71.03 | 67.43 | 3.48 | 3.44 | 3.13 | 2.17 | 0.32 |

Appendix II

## SUMMARY OF ALL REGRESSION EQUATIONS

Many combinations of independent ecologic variables were run in multiple regression analyses. The resulting equations, after the stepwise regression with backward elimination, are listed below. Many equations were repetitive after the elimination of insignificant variables and are only listed once. All of these equations were examined to determine the "best fit" equation for each cancer site for both sexes.

DEPENDENT VARIABLE: MALE LUNG

1. $Y = -.727 + .000011$ medinc $+ .0445$ educ $+ .0076$ urb02 $+ .0008$ popden $- .3052$ mfrg07 $+ .0089$ british + error

   $R^2 = .3899$ overall $F=4.15$ ($p < .01$)

2. $Y = 1.003 + .013$ lowinc $+ .0069$ educ $+ .001$ popden $- .247$ mfrg07 + error

   $R^2 = .3207$ overall $F=4.84$ ($p < .01$)

3. $Y = 1.003 + .013$ lowinc $+ .0069$ educ $+ .001$ popden $- .247$ mfrg07 + error

   $R^2 = .3208$ overall $F=4.84$ ($p < .01$)

4. $Y = 1.117 - .000001$ medinc $+ .0103$ educ $+ .0011$ popden $- .253$ mfrg07 + error

   $R^2 = .3145$ overall $F=4.70$ ($p < .01$)

5. $Y = 1.117 - ,00001$ medinc $+ .0161$ educ $+ .0008$ popden $- .084$ mfrg08 + error

   $R^2 = .2747$ overall $F=3.88$ ($p < .05$)

6. Y= .643 + .0207 lowinc + .0142 educ + .0007 popden -.074 mfrg08 + error
$R^2$ = .2848 overall F=4.08 (p < .05)

7. Y= 1.16 - .000003 medinc + .0106 educ + .0012 popden - .264 mfrg07 + error
$R^2$ = .3133 overall F+4.33 (p <.05)

DEPENDENT VARIABLE: FEMALE LUNG

1. Y= -.409 -.000002 medinc + .039 educ + .0082 urb02 + .0009 popden -.316 mfrg07 + .009 british + error
$R^2$ = .4211 overall F=4.73 (p <.001)

2. Y= -.469 + .007 lowinc + .037 educ + .008 urb02 + .0009 popden -.309 mfrg07 + .009 british + error
$R^2$ = .423 overall F=4.76 (p <.001)

3. Y= .9155 + .0225 lowinc + .0041 educ + .0011 popden - .24 mfrg07 + error
$R^2$ = .347 overall F=5.46 (p < .001)

4. Y= 1.42 - .00001 medinc + .0054 educ + .0012 popden - .266 mfrg07 + error
$R^2$ = .3363 overall F=5.19 (p <.01)

5. Y= 1.11 - .000007 medinc + .006 educ + .0009 popden - .04 seconemp + error
$R^2$ = .257 overall F= 3.55 (p < .05)

6. Y= .555 + .0298 lowinc + .011 educ + .0008 popden - .067

mfrg08 + error
$R^2$ = .3044 overall F=4.40 (p < .01)

7. Y= 1.13 - .00000029 medinc + .014 educ + .0012 popden - .324 mfrg07 - .0197 native + error
$R^2$ = .3791 overall F=4.52 (p < .01)

DEPENDENT VARIABLE: MALE STOMACH

1. Y= -2.44 + .00004 medinc + .074 educ + .010 urb02 + .001 popden -.355 mfrg07 + .016 british + error
$R^2$ = .5278 overall F=7.27 (p <.001)

2. Y= -1.07 -.018 lowinc + .06 educ + .010 urb02 + .0012 popden -.362 mfrg07 + .013 british + error
$R^2$ = .5116 overall F=6.81 (p < .001)

3. Y= .8504 + .005 lowinc + .013 educ + .0014 popden -.2539 mfrg07 + error
$R^2$ = .429 overall F=3.48 (p <.01)

4. Y= .86 + .00000066 medinc + .015 educ + .001 popden - .25 mfrg07 + error
$R^2$ = .3943 overall F=6.67 (p <.001)

5. Y= -2.21 + .00003 medinc + .065 educ + .0097 urb02 + .007 popden -.006 seconemp + .014 british + error
$R^2$ = .4714 overall F=5.80 (p <.001)

6. Y= -.98 -.016 lowinc +.053 educ + .010 urb02 + .0009 popden -.069 seconemp + .011 british + error
$R^2$ = .4585 overall F=.5.51 (p <.001)

7. Y= .945 - .000008 medinc + .0145 educ + .0015 popden -

.276 mfrg07 + error

$R^2$ = .3991 overall F=6.31 (p <.01)

DEPENDENT VARIABLE: FEMALE STOMACH

1. Y= .522 +.000023 medinc + .015 educ + .0015 popden -.354 mfrg07 + error

   $R^2$ = .2867 overall F=4.12 (p < .05)

2. Y= 1.22 - .0017 lowinc + .006 educ + .0017 popden -.383 mfrg07 + error

   $R^2$ = .2774 overall F= 3.94 (p <.0086)

3. Y= -.3.35 + .00006 medinc + .077 educ + .016 urb02 - .074 seconemp + .015 british + error

   $R^2$ = .297 overall F= 3.38 (p < .05)

4. Y= .028 -.0016 lowinc + .019 educ + .008 urb02 + .0009 popden -.062 seconemp + error

   $R^2$ = .2915 overall F=3.29 (p <.05)

5. Y= .546 + .00002 medinc + .018 educ + .0016 popden -.36 mfrg07 + error

   $R^2$ = .2846 overall F=3.78 (p <.01)

Appendix III

## Descriptive Statistics

This appendix section includes the descriptive statistics calculated by the SAS system for each of the 21 variables.

| Variable | N | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|
| MALELUNG | 47 | 1.0278936 | 0.3001788 | 0.6590000 | 2.2410000 |
| FEMLUNG | 47 | 1.0022128 | 0.2985711 | 0.4700000 | 2.1640000 |
| MALESTOM | 47 | 0.9265106 | 0.3344862 | 0.5040000 | 2.3910000 |
| FEMSTOM | 47 | 0.9218511 | 0.4840764 | 0.4690000 | 3.2230000 |
| URB01 | 43 | 56.3013953 | 25.2998963 | 13.0400000 | 100.0000000 |
| URB02 | 46 | 62.5408696 | 20.5249371 | 13.7000000 | 98.0600000 |
| POPDEN | 47 | 81.1446809 | 166.1648995 | 0.1000000 | 1001.60 |
| EDUC | 47 | 19.7444681 | 4.3174705 | 9.3300000 | 30.6900000 |
| LOWINC | 47 | 11.5446809 | 2.2607048 | 6.3000000 | 16.4000000 |
| MEDINC | 47 | 20388.79 | 3368.05 | 13020.00 | 29496.00 |
| MFRG07 | 47 | 1.3868085 | 0.4845846 | 0.5600000 | 2.4400000 |
| MFRG08 | 46 | 2.5423913 | 1.2866238 | 0.6800000 | 6.6700000 |
| PRIMEMP | 47 | 1.7538298 | 1.6659252 | 0.1600000 | 9.4000000 |
| SECONEMP | 47 | 4.3853191 | 1.9069972 | 1.4500000 | 8.8800000 |
| TEREMPL | 47 | 57.8453191 | 8.7314078 | 33.1600000 | 88.9000000 |
| BRITISH | 47 | 59.4738298 | 15.5738266 | 16.7300000 | 80.9600000 |
| FRENCH | 47 | 9.8219149 | 14.5095541 | 1.6100000 | 75.8300000 |
| DUTCH | 47 | 2.8719149 | 1.8201596 | 0.5300000 | 7.5800000 |
| GERMAN | 47 | 4.8717021 | 4.1582931 | 0.7300000 | 21.2600000 |
| ITALIAN | 47 | 2.1819149 | 2.5708234 | 0.1400000 | 9.4500000 |
| NATIVE | 47 | 2.3638298 | 5.8192535 | 0.1100000 | 30.6100000 |

## REFERENCES


Baum, A. and J. Singer. 1982. "Advances in Environmental Psychology" Environment and Health. 4: 237-267.

Blot, R., Fraumeni, J., Mason, T., and R. Hoover. 1979. "Developing Clues to Environmental Cancer: A Stepwise Approach with the Use of Cancer Mortality Data" Environmental Health Perspectives. 32: 53-58

Canada, Health and Welfare. 1988. The Smoking Behaviour of Canadians - 1986, Ottawa: Supply and Services Canada.

Devesa, S., Pollack, E., and J. Young. 1984. "Assessing Validity of Observed Cancer Incidence Trends" Journal of Epidemiology 119: 274-289.

Doll, R. and R. Peto, 1981. The Causes of Cancer, New York: Oxford University Press.

Hernberg, S. 1982. Prevention of Occupational Cancer International Symposium, series #46, Geneva.

Holland, W., Bennett, A., Cameron, I., Leeder, S., Schilling, R., Swan, A., and R.E. Waller. 1979. "Health Effects of Particulate Pollution: Re-Appraising the Evidence" American Journal of Epidemiology 110: 525-658.

Jones, K. and G. Moon, 1987. Health, Disease and Society, London: Routledge and Kegan Paul, chs. 2,3,4.

McLaughlin, J. and W. King, Cancer Incidence Projections for Ontario to the Year 2000. Toronto.

Morgenstern, H., 1982. "Uses of Ecologic Analysis in Epidemiologic Research" American Journal of Public Health. 72: 1336-1344.

Ontario Cancer Registry. Unpublished data. 1989.

Ontario, Ministry of the Environment, 1978. Regulatory and Economic Aspects of Air Pollution Control, Moncton.

Reynolds, D. 1989. The Geographic Distribution of Cancer Mortality in Ontario: An Ecologic Analysis on Ethnicity, Industry and SocioEconomic Characteristics 1976-1985. MSc Thesis for the Department of Preventative Medicine and Biostatistics.

Sebastiani, A., 1983. "The Causes of Cancer - Avoidable Risks Today", Recent Advances in Medicine 74(2): 131-138.

Shannon, H., Hertzman, J., Julian, J., Hayes, M., and N. Henry, 1988. "Lung Cancer and Air Pollution in an Industrial City - a Geographical Analysis" Canadian Journal of Public Health, 79(4): 255-259.

U.S. Department of Health and Human Services, 1985. The Health Consequences of Smoking: Cancer and Chronic Lung Disease in the Workplace. Report of the Surgeon General.

Ziegenfus, R.C., 1987. "Air Pollution and Health" Public Health and the Environment, In Greenberg, M.R.(ed.) Guilford Press, New York.