PORT COLBORNE

Protocol D

Cancer incidence and causes of mortality in a historical cohort of Port Colborne residents

19 September 2003

Protocol D: Cancer Incidence

Principal Investigator: **Paul J. Villeneuve, PhD**

Co-Investigator: **Edward M. Sellers, MD, PhD**

Ventana Clinical Research Corporation
340 College Street, Suite 400
Toronto, ON
M5T 3A9

19 September 2003
Version 4.1

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

PROTOCOL SYNOPSIS

TITLE	Cancer incidence and causes of mortality in a historical cohort of Port Colborne residents
SPONSOR	INCO Ltd.
STUDY SITE	Toronto, Ontario
PRIMARY OBJECTIVE(S)	1. To determine whether adults (≥20 years of age) who were resident in Port Colborne for at least one year between 1982 to 2000 have increased incidence and mortality rates of respiratory cancer relative to a sample of residents selected from a group of similar Ontario communities, and 2. To evaluate variations in this risk according to length of residency.
SECONDARY OBJECTIVES	To investigate adults (≥20 years of age), who were resident in Port Colborne for at least one year between 1982 and 2000, in order to: 1. Determine whether they have increased incidence and mortality rates of respiratory cancer relative to the Ontario population 2. Describe the incidence of all cancers (excluding skin), in total and by anatomical site, in comparison to a sample of residents selected from a group of similar Ontario communities and to the general population of Ontario 3. Describe the mortality patterns in comparison to a sample of residents selected from a group of similar Ontario communities and to the general population of Ontario 4. Explore the distribution of incident cases of adult respiratory cancer in relation to the potential for differential exposure to the CoCs within Port Colborne. Based on the above analyses, a further secondary objective will be to describe the role that confounding variables and residential mobility may have had on risk estimates generated in previous ecological studies of cancer incidence in Port Colborne conducted by the Niagara Department of Health.
DESIGN AND METHODOLOGY	A cohort of Port Colborne residents and a sample of residents of selected comparator communities will be linked to Canadian cancer incidence and mortality data. Exposure to chemicals of concern (CoCs) will be inferred from residential histories of the cohort members available on an annual basis between 1982 and 2000. Using the assembled cohort, rates of respiratory cancer among Port Colborne adult residents will be compared to a sample of residents of the comparator communities and to Ontario as a whole. Respiratory cancer rates will also be compared between Port Colborne cohort members according to estimated levels of exposure to certain CoCs. Cohort members will be identified from tax filing data contained within the Statistics Canada Annual Estimates for Families and Individuals File (T1FF) for the years 1982 to 2000. Secondary analyses will compare incidence rates between Port Colborne and the two referent groups for all cancers and mortality for selected underlying causes. Various methods will be used to control for, or assess, the impact of important confounders such as income and smoking. Furthermore, individuals that were employed by INCO will be identified to better assess health risks resulting from environmental exposures to the CoC by controlling for occupational exposures. Individuals with an employment history with INCO will be identified from electronic files provided to Statistics Canada that contain

	surname, given name, birth date and years of employment. Exposure information prior to 1982 will not be available for cohort members from the T1FF. Sensitivity and sub-analyses will be performed to assess the impact of missing residential history data on the relative risk estimates.
POPULATION	Residents of Port Colborne and Ontario over the period 1982 to 2000, as identified through tax filing data
SUBJECT PARTICIPATION	There will be no direct participation of Port Colborne or other Ontario residents, as the occurrence of cancer incidence and mortality will be ascertained using existing population-based databases. Analysis files will be stripped of personal identifying information so that no individual can be identified.
OUTCOME MEASURES	*For primary objectives:* Incidence and mortality rates of respiratory (i.e. lung, tracheal, bronchal, nasal and laryngeal) cancer among residents of Port Colborne, and among a sample of residents from a group of comparator communities. *For secondary objectives:* Incidence rates for site-specific cancer (excluding skin) and causes of mortality among residents of Port Colborne, among a sample of residents from a group of comparator communities and the province of Ontario. Incidence rates of respiratory cancer will also be examined for Ontario.
DATA COLLECTION & ANALYSES	For each cancer outcome, the number of incident events will be determined for strata defined by age-group, community, sex, income, calendar year, previous employment status with INCO, and estimated exposure to the CoCs based on residential history. Similar cross-classification tables will be constructed for mortality outcomes. These cases will be identified from the Canadian Cancer Database and the Vital Statistics Death Database for the period from 1982 to 2000. Person-years of follow-up within the cohort will also be estimated for each stratum. Exposure will be defined by a cumulative time-dependent covariate constructed from the residential histories of each cohort member that are available on an annual basis from postal code information in the T1FF file. The Ontario comparator communities have been chosen such that they are similar to Port Colborne with respect to several sociodemographic characteristics. Poisson regression will be used to estimate series of rate ratios for cancer incidence and for mortality. Specifically, disease rates will be compared relative to those of the comparator communities (combined). For comparison to the Ontario rates, the expected number of cases will be estimated by multiplying the Ontario rates by the accumulated age-sex specific person-years in the cohort. The standardized incidence (or mortality ratio) will then be calculated by dividing the observed number of disease outcomes by the number estimated using Ontario rates. For the standardized rate ratios (i.e. the SMR and SIR) the confidence intervals will be calculated and used to assess whether there are statistically significant differences between these populations. Poisson regression methods will be used to perform an internal comparison of respiratory cancer incidence rates across Port Colborne residents based on their estimated residential CoC exposure levels with or without occupational exposure. Steps will be taken to ensure the accuracy of the estimated standard errors obtained from Poisson regression modelling. Specifically, where relevant, appropriate methods will be applied to correct for correlations in the data (overdispersion), such that the precision of the risk estimates are not overstated. Because information on age, sex and income are available, we will be able to adjust the rate ratios

for these characteristics. Moreover, we will evaluate the impact that differential community smoking rates may have contributed to observed differences in rates of cancer incidence or mortality. This will be done by modelling smoking prevalence data obtained from the 1985 and 1994 Ontario Health Surveys. We will also evaluate the role smoking may have had on our findings by comparing risk estimates across a series of smoking and non-smoking related health conditions, by conducting sensitivity analyses and by adjusting for other variables that are recognized correlates of smoking status.

ABBREVIATIONS

ASIR	Age-Standardized Incidence Rate
ATSDR	American Agency for Toxic Substances and Disease Registry
CCR	Canadian Cancer Registry
CDI	Chronic Daily Intake
CHAP	Community Health Assessment Project
CI	Confidence Interval
CoCs	Chemicals of Concern
CSTEE	Scientific Committee for Toxicity, Ecotoxicity and the Environment
EPA	Environmental Protection Agency
FSA	Forward Sortation Area
GEE	Generalized Estimating Equations
GST	Canadian Federal Government Goods and Services Tax
IARC	International Agency for Research on Cancer
ICD	International Classification of Disease
ICNCM	International Committee on Nickel Carcinogenesis in Man
INCO	INCO, Ltd., Port Colborne
JWEL	Jacques Whitford Environmental Consultants
MOE	Ministry of the Environment (Ontario)
NCIRS	National Cancer Incidence Reporting System
PPM	Parts per million
PTCR	Provincial and Territorial Cancer Registries
SES	Socioeconomic Status
SIN	Social Insurance Number
SIR	Standardized Incidence Ratio
SMR	Standardized Mortality Ratio
T1FF	Statistics Canada Annual Estimates for Families and Individuals File
USDHHS	U.S. Department of Health and Human Services
WHO	World Health Organization

1. RESEARCH OBJECTIVES

The purpose of this study is to examine patterns of cancer incidence and investigate the causes of mortality among residents of Port Colborne from 1982 to 2000. Motivating the proposed research are the high levels of nickel, arsenic, cobalt and copper (the Chemicals of Concern or CoCs) that are present in the soils in some areas of Port Colborne. These soil contamination levels provide a record of the historical environmental contamination in the community resulting from metal refining operations.

The CoCs are the focus of the study protocols that form the Community Health Assessment Project (CHAP). As described in Section 2.3 Literature Review, the scientific literature provides evidence of an association between the development of respiratory cancers with inhalation exposure at certain doses to nickel, arsenic and possibly cobalt. Ingestion of high levels of arsenic has been linked with other types of cancer.

Possible associations between exposure to the CoCs and cancer continue to be a major concern for some Port Colborne residents who feel their health concerns have not been satisfactorily addressed by previous research. Patterns of mortality in Port Colborne are also of interest to the community and are included as outcomes under investigation. Their inclusion also provides an opportunity to evaluate potential biases in the assessment of risk for cancer outcomes. The specific objectives of the study are as follows:

Primary objectives

1. To determine whether adults (≥20 years of age), who were resident in Port Colborne for at least one year between 1982 and 2000, as defined in the T1FF, have increased incidence and mortality rates of respiratory cancer relative to a sample of residents of a representative series of Ontario communities.
2. To examine variations in the risk for respiratory cancer among Port Colborne residents relative to a sample of residents from a representative series of Ontario communities according to duration of residency in Port Colborne.

Secondary objectives

1. To determine whether adults (≥20 years of age), who were resident in Port Colborne for at least one year between 1982 and 2000, as defined in the T1FF, have increased incidence and mortality rates of respiratory cancer relative to the Ontario population.
2. To describe the incidence of all cancers (excluding skin), in total and by anatomical site, of adults residing in Port Colborne for at least one year between 1982 and 2000, as defined in the T1FF, in comparison to a sample of residents from a representative series of Ontario communities and to the general population of Ontario.
3. To describe the mortality patterns of adults residing in Port Colborne for at least one year between 1982 and 2000, as defined in the T1FF, in comparison to a sample of residents from a representative series of Ontario communities and to the general population of Ontario.

4. To explore the distribution of incident cases of adult respiratory cancer among Port Colborne residents over the period from 1982 to 2000, as defined in the T1FF, in relation to the potential for differential exposure to the CoCs within Port Colborne.
5. To explore the role that confounding variables and residential mobility may have had on previous ecologic studies of cancer incidence in Port Colborne conducted by the Regional Niagara Public Health Department.

2. RATIONALE

From a health perspective, preliminary inquiries in Port Colborne revealed that residents identified cancer as one of their most serious health concerns related to the chemical contamination in their community (Ventana CRC, 2001). Substantiating the community concerns are the observed associations between cancer and the CoCs reported in several epidemiologic and laboratory studies. This existing body of research has demonstrated increased risks of respiratory cancer from exposure to nickel, arsenic and possibly cobalt in occupational settings where workers have been chronically exposed to high levels of these metals. Several studies have also found increased risks for lung cancer with environmental exposures to airborne arsenic, and to skin, bladder, kidney, prostate and liver cancers with exposure to high levels of arsenic in drinking water. The findings from animal and cellular studies suggest the CoCs have carcinogenic effects across a range of concentrations and durations of exposure. However, with the exception of arsenic exposures, there have been few epidemiologic studies that have evaluated the relationship between non-occupational exposure to these metals and the occurrence of cancer.

In response to community concerns, the Regional Niagara Public Health Department conducted two ecologic investigations of cancer incidence in Port Colborne from 1979 through 1996 (Health Services Department Region of Niagara, 1997; MOE, 2000). These cancer incidence studies were conducted as part of the health department's assessment of the potential health risks of the reported soil levels of nickel, cobalt and copper in the community. Port Colborne cancer incidence rates for females from 1982 to 1991, and for males for the periods of 1984 to 1988 and 1989 to 1991, were not significantly different from expected rates calculated from the appropriate Ontario-wide incidence data. In the five-year period between 1979 and 1983, the incidence of lung cancer in Port Colborne males was elevated (SIR=1.35, 95% CI=1.03-1.72) relative to Ontario rates, the only significant finding in these ecologic analyses. This prompted a further analysis, which was conducted in 2000 using cancer incidence data from 1987 through 1996. When compared to the Ontario general population, no significantly increased rates of cancer incidence were found in Port Colborne (MOE, 2000).

Although this cancer incidence investigation rapidly yielded some information regarding cancer incidence in Port Colborne, the limitations of the study design also affected the conclusions that could be drawn from these results. The ecologic design compares incidence rates using aggregate data; as a result, cancer incidence data were not linked to residence history at the individual level. Because comparisons were made based on

residency at the time of cancer diagnosis, with no additional residential information available, no control could be made for the effects of population mobility on the risk estimates. Therefore, if long-time residents of Port Colborne moved out of the city and were subsequently diagnosed with cancer, those instances of cancer would be incorrectly captured in the data for the referent group, or not captured at all had they moved outside of the province. Conversely, individuals who had resided elsewhere and had recently moved into Port Colborne would be included in the calculation of cancer rates for Port Colborne. More importantly, no distinction could be made based on how long individuals had lived in Port Colborne.

While the published risk estimates were adjusted for age and sex, the study was unable to account for the role of other risk factors that may have biased (confounded) comparisons of cancer incidence rates between Port Colborne and Ontario. As cigarette smoking accounts for the majority of all cases of respiratory cancer (USDHHS, 1989), any observed population rate differences could be explained by differences in this risk factor. In addition, this study focused solely on cancer incidence, and did not explore any patterns of mortality in the Port Colborne community. Therefore, it was unable to address community concerns regarding any possible associations between the CoCs and mortality.

Protocol D is part of the CHAP, a comprehensive series of integrated studies carefully designed to address the public's health concerns in Port Colborne. This protocol has been developed to build on the previous ecologic study conducted by the Regional Niagara Public Health Department, and to avoid many of the limitations of that study in an effort to further and more clearly understand cancer incidence rates and mortality patterns in Port Colborne.

We propose to build a historical cohort of Port Colborne residents using an administrative database and subsequently link this cohort to existing cancer and death registry data, a design that has four important strengths:

1. Ecologic bias occurs when individual-level influences are masked in a group analysis setting. In this study, information for each member of the cohort will be available on an individual level, thereby reducing the potential for ecologic bias.
2. Exposure will be based on residential history, and will include annual follow-up of all cohort members who meet minimum Port Colborne or comparator community residency requirements. This follow-up occurs over an extended time span (1982 to 2000). This will reduce the potential for bias due to population mobility, and will permit an assessment of the effects of such mobility on observed cancer incidence and mortality rates.
3. Some sociodemographic information will be available at a regional level, and may prove useful to control for the associated confounding role of this determinant of health.
4. We will be able to identify those Port Colborne cohort members with previous employment at INCO; therefore, our analyses will be able to examine the relationship between environmental (i.e. residential) exposures to the CoCs while controlling for potential occupational exposures. In so doing, different sources of potential exposure

in the Port Colborne community will be separated, thereby avoiding a dilution effect in within-community comparisons that does not take occupational exposures into account.

Indirect methods for dealing with potential confounding by smoking status have also been included in the study design. As will be detailed later in this protocol, these methods include: (1) comparison of risk estimates across health outcomes that are widely recognized to be affected by smoking relative to non-smoking related conditions, (2) control for smoking-related variables, (3) use of population-based survey data in Ontario and (4) sensitivity analyses.

The proposed study has some important limitations (see below), in particular a limited scope for attribution of causality in the relationship between the CoCs and cancer or mortality. However, we believe this is the strongest design feasible in the current context. Despite the limitations, this work will provide additional and valuable insights to both the issue of whether Port Colborne residents have higher rates of cancer relative to other comparator communities, and whether potential associations exist between environmental exposure to the CoCs and cancer incidence and mortality in Port Colborne.

2.1 Study Limitations

An assessment of epidemiological study designs and data sources was carried out in order to determine the most suitable database and design for addressing the objectives of Protocol D (Ventana CRC, 2003a). Both internal and external scientific reviews determined that the historical cohort design and Statistics Canada's Annual Estimates for Families and Individuals File (T1FF) would be best suited for evaluating the objectives and for providing information in a timely manner. Although the historical cohort study has a number of advantages over other study designs for answering community concerns, there are limitations related to both design and data availability.

The most challenging issue of the study is to assign individual level of exposure to the cohort members. Our assignment of these exposures is limited by our inability to track the residency of cohort members from birth and therefore construct a lifetime profile of exposure for each individual. Data in the T1FF are only available from 1982 to 2000; this may introduce exposure misclassification given that a subject's residency status is only known for a portion of his/her lifetime. Using the T1FF alone, it is not possible to track exposures prior to 1982. This is particularly relevant for environmental causes of cancer that have a long induction period. Hertz-Picciotto (1998) note that past exposures or residences are relevant for studying diseases with long latency periods, such as cancer or diseases caused by long-term chronic insults. Assuming that this exposure misclassification is non-differential according to the incidence of disease, any real difference would be underestimated. For cohort members with cancer and mortality outcomes identified through record linkage, place of birth information can be extracted from both the cancer and mortality database, thus yielding some information about residential history prior to 1982. Census data also provides summary data to describe the

mobility patterns of communities, thereby providing useful information to conduct sensitivity analyses to evaluate the magnitude of potential bias resulting from this source of exposure measurement error.

Residential location is used as a proxy measure of potential exposure. Researchers in environmental epidemiology have noted, however, that residential location or subject characterization of potential exposure based on environmental data is useful when the exposures are prevalent in some, but not all, geographic areas or time periods under study (Hertz-Picciotto, 1998). This is the case with a Port Colborne versus comparator community comparison. Our study will use three exposure indices: residency in Port Colborne (Yes or No), proximity of residence in relation to the INCO facility (in kilometres) and measures of CoCs taken from soil samples in Port Colborne (in ppm). Misclassification of exposure may be larger amongst Port Colborne cohort members given that we are not fully able to account for differences in individual levels of exposure using direct measures.

For several cancer sites, the relatively small population size of Port Colborne precludes meaningful comparisons between Port Colborne and the comparator communities and comparisons within Port Colborne itself. The primary study objective, however, is to compare incidence and mortality rates of respiratory cancers in the Port Colborne population to other populations, as these cancers have been linked to CoC exposures in occupational settings. Cancer incidence rates offer a distinct advantage over mortality rates, since they are not subject to differences in treatment and management for this malignancy that may occur between regions. However, given that individuals diagnosed with respiratory cancers have a poor prognosis, we expect no appreciable bias in risk estimates for respiratory malignancies when mortality data rather than incidence are modelled. For both mortality and incidence, respiratory cancer outcomes are sufficiently common, and thus provide the study with the necessary statistical power to perform this external comparison and therefore examine this objective (see Section 5 for additional details).

Finally, smoking is the most important risk factor for respiratory cancer, and plays an important role in the development of a number of other health conditions. This study will have no direct measures of active or passive smoke exposure available for each cohort member. However, a number of methodological strategies are incorporated into this study in order to assess the impact of this important confounder. These are described in greater detail in Section 4.3.2.

2.2 Objectives of the CHAP Research

To place the current study in the broader context of the proposed CHAP studies, a summary of the project follows. Higher than background levels of the four CoCs (nickel, arsenic, cobalt and copper) have been observed in extensive soil sampling conducted in the Port Colborne area, and have been attributed to historical emissions from metal refinery operations. The primary research objectives of the CHAP are to:

1. Determine whether the health of the Port Colborne community varies significantly from that in samples from the population of Ontario, and
2. Understand the relationship between the health of the community and its potential environmental exposure to the CoCs.

To address these objectives, four critical study areas were identified, as outlined in the *Overview of Proposed CHAP Research in Port Colborne* (Ventana CRC, 2002). These study areas represent converging health assessment strategies (including general and comprehensive health questionnaires, medical testing and the review of existing health registries) that are outlined in Protocols A, C and D, and will be subsequently developed in Protocol B, if warranted. The four proposed studies are listed below; the specific research objectives of each individual study are outlined within each of the corresponding protocols.

Study A: *A self-reported health assessment of the Port Colborne community – Phase I of risk assessment of current Port Colborne residents*

Study B: *Case-control study(ies) of selected health conditions using a sample of Port Colborne residents - Phase II of risk assessment of current Port Colborne residents [If warranted]*

Study C: *Hospital discharge patterns among Port Colborne residents: A comparative analysis to Ontario rates*

Study D: *Cancer incidence and causes of mortality among a historical cohort of Port Colborne residents*

2.3 Literature Review

2.3.1 Relationship between CoCs and human health

Information about the relationships between the CoCs and cancer or other causes of mortality is based primarily on epidemiologic studies of occupational exposures and on *in vivo* (animal) and *in vitro* (cellular) studies. It should be noted that prolonged human exposure to any single CoC at higher than background levels is rare. Therefore, most epidemiologic studies of CoC effects include exposures to a number of other chemicals and metals, which cannot be evaluated separately, and may modify the impact of the exposure (IARC, 1991; Sabbioni et al., 1994; Hayes, 1997). In addition, further scientific study is required to differentiate the relative importance of effects of exposure to different metal species and their compounds (Hayes, 1997). At the same time, "metals share certain physical and chemical features, and it is reasonable to speculate that common mechanisms for carcinogenicity may operate" (Hayes, 1997).

Most of the epidemiology surrounding the health outcomes associated with these metals and metal compounds is concerned with inhalation exposures. There is also some discussion of oral and dermal exposures, particularly involving animal toxicology studies. Comprehensive literature reviews by the International Agency for Research on Cancer (IARC) and the American Agency for Toxic Substances and Disease Registry (ATSDR)

were incorporated into this review, along with a number of individual epidemiologic studies and other review papers. The following four sections summarize the current understanding of the effects of each CoC on human mortality and cancer incidence.

Nickel

Nickel compounds are recognized by the IARC as human carcinogens, primarily in the upper respiratory tract (nasal) and lung (IARC, 1990). This is based in part on epidemiologic studies that have observed increased risks of cancer among workers with high exposure to nickel compounds from refinery dust. In 1990, the International Committee on Nickel Carcinogenesis in Man (ICNCM) published findings of a study consisting of 10 cohorts among workers in nickel mining, smelting, refining and specialty use (Doll et al., 1990). The largest cohort in this report included approximately 54,000 workers at INCO nickel, smelting and refining facilities in Sudbury and Port Colborne, Ontario. Increased risks of lung and nasal cancer were observed among workers in sinter plant operations (leaching, calcining, sintering) where oxidic and sulfidic nickel compounds were the primary exposures from nickel refinery dust (≥10 mg nickel/m^3). A higher incidence of lung and nasal cancer was observed for workers exposed to both soluble and less-soluble nickel compounds, compared to those exposed to less-soluble nickel compounds alone. This finding indicated an effect of soluble nickel, or an interaction between soluble and less-soluble nickel compounds (>1 mg nickel/m^3), on the incidence of lung and nasal cancer. Although the standardized incidence ratio (SIR) for laryngeal cancer was not significant in the Doll et al. study, a more recent incidence study (Julian and Muir, 1996) reports significant findings. The study of laryngeal cancer and occupational exposures among INCO and Falconbridge workers in Sudbury and Port Colborne reported a significant elevated incidence (relative to the Ontario population) for Ontario mill workers with greater than 25 years of exposure (SIR=4.00 (95% CI 1.47-8.71)) (Julian and Muir, 1996).

Cellular studies have provided further evidence that implicates nickel as a carcinogen. Experimental studies have found that nickel produces DNA strand breaks, DNA-protein links and inhibits DNA repair (Hayes, 1997). Respiratory exposure to soluble nickel monoxide, nickel oxide and nickel subsulfide has induced malignant tumours of the lung in rodents. Various durations and exposures have resulted in rat and/or mouse tumours indicating species-specific effects (ATSDR, 1997).

The World Health Organization (WHO) also recognizes that nickel compounds are human carcinogens by inhalation exposure based on the studies of occupationally exposed workers and evidence reported from animal studies. Further, the WHO assumes a linear dose-response relationship, and does not recommend any safe level of nickel exposure (WHO, 2000). However, there are few studies that investigate lower (non-occupational) levels of nickel exposure, in particular those linking nickel uptake from the environment and cancer incidence in the general population. In a study of the geographical distribution of respiratory cancer in New Caledonia, a French territory in the South Pacific, a significant excess of primary lung cancer was observed to be associated with a greater

number of stays in, and with more person-years of exposure to, areas classified as mining zones as compared to other areas (Leclerc et al., 1987). The observed relationships may, however, be due to uncontrolled factors such as smoking. More recently, a case-control study of lung cancer was undertaken in the population of Karpinsk, a small community in northern Russia with unusually high cancer morbidity rates (Katsnelson et al., 2002). The authors state that unspecified "carcinogenic metals" in the soil around homes were examined as potential risk factors, but their influence could not be demonstrated.

"Studies in both humans and animals indicate that the respiratory system is the primary target of nickel toxicity following inhalation" (ATSDR, 1997). The ATSDR based its assessment of the mortality effects of nickel in humans on epidemiologic studies of occupational health. Two studies reported an increase in non-malignant respiratory causes of death among nickel-exposed workers who were concomitantly exposed to various other metals. Another five occupational studies reported that no increases in death due to respiratory conditions were observed. No other non-cancer effects of nickel were found in the review of the literature (ATSDR, 1997).

Cobalt

Cobalt and cobalt compounds have been classified by the IARC as possibly carcinogenic to humans (Group 2B). This classification was based mainly on evidence from animal experiments with cobalt compounds, some of which have shown significant increases in respiratory tumour formation (IARC, 1991; ATSDR, 2001). However, extrapolation of effects from one species to another is problematic, and a number of the animal studies reviewed were deemed insufficiently detailed to provide strong evidence (IARC, 1991). The human evidence was deemed by IARC to be inadequate (IARC, 1991), a summary consistent with the assessment by the ATSDR that "cobalt [metal] has not been shown to cause cancer in humans by the inhalation, oral or dermal exposure routes." Occupational exposure to hard metal (a metal alloy with a tungsten carbide and cobalt matrix) has been demonstrated to increase the risk of lung cancer.

Information about cancer risks from occupational cohorts exposed to cobalt is limited, with a small number of studies and most exposed populations being concurrently exposed to other metals and dusts. Six reports were identified that examined the relationship between cobalt exposure and lung cancer in five separate study populations. In the three studies that examined mortality among hard-metal workers, excess risks were found for lung cancer mortality, but not all-cause mortality (Hogstedt and Alexandersson, 1990; Lasfargues et al., 1994; Moulin et al., 1998). These risks (standardized mortality ratio (SMR) (95% CI) = 1.34(0.77-2.13); 2.13(1.02-3.93); 1.30(1.00-1.66), respectively) remained after adjusting for smoking, and showed some association with various measures of increased exposure.

Two other reports were based on a group of workers in a French electrochemical plant. The initial report on the cohort found an excess risk of lung cancer (SMR=4.66; 95% CI 1.46-10.64) in workers exposed to cobalt metal particles (Mur et al., 1987), whereas the second report, which included additional follow up time for the original cohort, showed no

excess risk (SMR=0.85; 95% CI 0.18-2.50) (Moulin et al., 1993). The main explanation for the difference in the results appears to be the use of different case-ascertainment strategies for the two analyses (Lison et al., 2001), leaving the utility of the reports in some question. Women working in two Danish porcelain factories and exposed to a mineral cobalt compound were observed to have a small excess lung cancer risk as compared to a standard population and to unexposed controls. However, the unexposed controls also showed a small increased risk for lung cancer when compared to the standard population (Tuchsen et al., 1996). No epidemiologic studies showing increased risks for other cancer sites were found, nor did we find any such studies that evaluated non-occupational cobalt exposures.

Cobalt has been found to produce adverse non-cancer effects in humans by the inhalation, oral and dermal routes (ATSDR, 2001). Chronic cobalt exposure may affect various organ systems, primarily the lungs, although cardiovascular, renal, hepatic and ocular effects have also been noted. Occupational exposure to cobalt and tungsten carbide (hard metal) has been linked to "hard metal asthma" and "hard metal disease" characterized by respiratory problems of variable severity. Prolonged cobalt exposure has also been linked to altered levels of thyroid hormones, suggesting an effect on thyroid metabolism in those exposed in an occupational setting (Prescott et al., 1992; Swennen et al., 1993). No epidemiologic study of the relationship between cobalt and non-malignant causes of death was found.

Copper

There is scant information regarding the effects of environmental exposure to copper on cancer risk, with no epidemiologic evidence available for assessing the human carcinogenicity of the metal. In a 1990 toxicological profile for copper, after stating that "no studies were located regarding carcinogenic effects in humans" for any exposure route, the ATSDR concluded that "an elevated incidence of cancer has not been observed in humans or animals exposed to copper via inhalation, oral or dermal routes of exposure" (ATSDR, 1990). Furthermore, copper is not classified as either an animal or human carcinogen by the IARC. No other studies were identified which examined the effects of occupational or environmental copper exposure on the risk for cancer.

The ATSDR also states that there is little information on copper toxicity in man. Acute ingestion of large doses of copper may result in nausea and vomiting, liver and kidney damage, and adverse effects on the blood. However, such exposures occur rarely, since copper can be tasted at levels well below those required to produce toxic effects. Furthermore, the body has efficient mechanisms for blocking absorption of excess ingested copper, thereby making chronic oral exposure less of a concern from a health perspective. Long-term inhalation exposure can result in acute respiratory irritation, dizziness, headaches and diarrhea (ATSDR, 1990).

Arsenic

Arsenic and arsenic compounds are classified by the IARC as human carcinogens (Group 1) (IARC, 1980; IARC, 1987; Hayes, 1997). There is substantial evidence that inorganic arsenic causes cancer in humans, through both respiratory exposure and ingestion in drinking water and other sources (Cantor, 1997; Hayes, 1997; ATSDR, 2000).

Occupational studies have clearly demonstrated both associations and dose-response relationships between airborne arsenic exposure levels and increased risk for lung cancer (Hayes, 1997; Lubin et al., 2000; ATSDR, 2000). In addition, there have been reports of intestinal, stomach, colon, bladder, prostate and bone cancers associated with inhalation exposure of arsenic (Enterline et al., 1995; Bulbulyan et al., 1996; Wingren and Axelson, 1993; Lubin et al., 2000). However, this evidence is much weaker than that for lung cancer, with some of the reports showing only marginally or non-statistically significant associations (Bulbulyan et al., 1996; Lubin et al., 2000). These studies include those with low power due to small numbers of cases (Enterline et al., 1995; Bulbulyan et al., 1996) and studies where the relationship may have been confounded by other exposures (Bulbulyan et al., 1996; Wingren and Axelson, 1993).

We found eight studies that investigated the effects of living in proximity to arsenic-emitting smelters and other industrial arsenic sources (Frost et al., 1987; Rom et al., 1982; Cordier et al., 1983; Pershagen, 1985; Blot and Fraumeni, 1975; Xu et al., 1989; Matanoski et al., 1981; Brown et al., 1984). Seven of the studies showed evidence of increased lung cancer risks in populations living closer to the arsenic-emitting sources. These studies included both standardized mortality ratio analyses and case-control designs, and were conducted in Canada (Quebec), Sweden, the United States of America and China. The eighth study, from El Paso, Texas, did not show any increased risk for lung cancer, as compared to breast and prostate cancer controls (Rom et al., 1982).

There is convincing evidence from a number of epidemiologic studies that exposure to inorganic arsenic in drinking water increases the risk of skin cancer (ATSDR, 2000; Cantor, 1997). A number of large-scale epidemiologic studies have also detected increased risk for bladder, kidney, liver, lung and prostate cancers associated with arsenic ingestion (ATSDR, 2000). Most of the studies of arsenic ingestion have been conducted in areas with naturally occurring high levels of arsenic in the water, whereas the effects of industrial emissions have been less well documented. One study of a U.S. cohort whose members lived in the vicinity of several industrial arsenic sources did not observe any increase in skin cancer incidence rates relative to the general population (Wong et al., 1992).

Epidemiologic studies have also linked inhalation of inorganic arsenic to non-malignant causes of death. Copper smelter workers in Montana were observed to have significantly increased risks for all causes of death, all cancers, diseases of the nervous system and sense organs, non-malignant respiratory diseases and emphysema (Lubin et al., 2000). Other studies showed some indications of excess risks of non-malignant respiratory mortality associated with arsenic exposure, but none were deemed conclusive (ATSDR,

2000). Similarly, several studies have been suggestive of cardiovascular effects of arsenical inhalation exposures (Enterline et al., 1995; Welch et al., 1982; Wall, 1980; Tollestrup et al., 1995; Qiao et al., 1997), but these findings have not been confirmed elsewhere or in further follow-up of the original cohorts (ATSDR, 2000; Lubin et al., 2000; Wall, 1980; Järup et al., 1989).

2.3.2 Port Colborne exposure to CoCs

The geographical extent of the CoC contamination of Port Colborne soils was determined by the Ontario Ministry of the Environment through extensive soil sampling in the area (MOE, 2000). Contour maps were developed based on the sample data from 1998 and 1999 soil survey sites, and historical soil survey data from 48 sites dating back to 1990. The resultant soil contamination contour maps developed for each of the CoCs demonstrate significant variation in the potential for exposure to the CoCs across the Port Colborne community. Contours range from 200 to 1000 ppm nickel and copper, 50 to 150 ppm cobalt and 25 to 45 ppm arsenic, in the upper five centimetres of soil (Appendix I). These maps provide an estimate of the area in and around the city of Port Colborne that has been impacted by decades of INCO emissions and atmospheric deposition. This is concluded from an analysis (performed by Jacques Whitford Environmental Consultants (JWEL)) that includes the sampled soil metal concentrations, distance from the source, and prevailing wind directions (JWEL, 2001).

As evidenced from the pattern of soil metal contamination, increased levels of soil nickel are the most widespread throughout the community. Arsenic, cobalt and copper are essentially localized to a small mostly non-residential area east, northeast of the INCO plant (Appendix I). These CoC soil levels are indicative of contamination levels that exceed the Ministry of Environment's phytotoxic generic effects based soil guideline levels (i.e. 200 µg/g for nickel, 25 µg/g for arsenic, 300 µg/g for copper and 50 µg/g for cobalt) (MOE, 1997). Canadian communities that are not exposed to anthropogenic sources of metals from area smelters or sewage sludge disposal do not exhibit comparable soil CoC levels (Canadian Council of Ministers of the Environment, 1999). Soil CoC levels measured in uncontaminated areas are much lower, with phytotoxic generic background based soil guideline levels set at 43 µg/g for nickel, for example (MOE, 1997).

The pollution emission profile within Port Colborne has changed historically since INCO first began operations in 1918. The level of fugitive emissions shows an increase in nickel emissions from 1918 to 1960 with the highest rates of deposition occurring during the 1940s and 1950s. Nickel emissions during the 1940s and 1950s (~500 tonnes/yr) were almost twice as much as that observed during the two decades previous to 1940 (~300 tonnes/yr). Pollution control measures instituted during the 1960s dramatically decreased nickel emission rates to the levels observed until the 1980s (<50 tonnes/yr) (JWEL, 2001). Following this period, only marginal fugitive emissions could be observed up to the time the process emission stacks were demolished in 1995. Port Colborne residents have

therefore been potentially exposed over time to the CoCs through inhalation of polluted ambient air and inhalation/ingestion of contaminated soils.

Two extensive environmental risk assessment reports were produced by the MOE for Port Colborne (MOE, 2000) and for the Rodney Street community (MOE, 2002), in particular. These reports were generated in light of the observed increased soil CoC levels and described the potential for adverse human health effects from estimated exposures to CoCs based on inhalation, ingestion and dermal exposure routes. For example, the plausible "worst case" exposure model indicates that when the Rodney Street community nickel exposure from all sources is averaged over a lifetime, the resulting chronic daily intake (CDI) estimate (all age groups) is about 8 µg/kg/day or 40 per cent of the United States Environmental Protection Agency's (US EPA) reference dose (R*fD*) of 20 µg/kg/day. The US EPA lifetime averaged exposure reference dose is defined as the dose below which exposure averaged over a lifetime is unlikely to result in adverse health effects.

An individual's exposure to environmental contaminants is a function of characteristics such as age, gender and the duration and intensity of exposure. Both MOE reports concluded that the calculated risks to residents of Port Colborne and Rodney Street were extremely low or nonexistent from the current measured levels of CoCs in the soil. Although a sophisticated exposure assessment process was used, it should be noted that the study did not simultaneously measure both exposure and health outcomes. The US EPA has warned that environmental risk assessment techniques cannot be validly used to accurately predict the incidence of human disease or the type of effects that chemical exposures have on humans (US EPA, 1986).

The Rodney Street risk assessment used air-monitoring data from several sources and locations to estimate recent nickel concentrations in the air. These data were obtained from the MOE sampling station just north of Rodney Street, which operated between 1992 and 1996, and air sampling done during the summer of 2000 near schoolyards in Port Colborne. The highest annual average nickel concentration in ambient air in Port Colborne for risk assessment purposes was estimated to be 33 ng/m^3 (MOE, 2001). Since nickel emission rates were 400 to 500 times higher during the 1940s and 1950s as compared to the recent past (JWEL, 2001), nickel concentrations in ambient air during these earlier periods were undoubtedly much higher. Atmospheric concentration of nickel in industrialized areas has been estimated to be in the range of 120 to 170 ng/m^3 (Norseth and Piscator, 1979). In comparison, nickel air concentrations in occupational settings with observed increased lung and nasal cancers were much higher, ranging from >1 to ≥10 mg/m^3 (see Section 2.3.1).

Although there is this evidence for nickel carcinogenicity in an occupational setting, in 2001 the Scientific Committee for Toxicity, Ecotoxicity and the Environment (CSTEE) issued an opinion of the human health risks posed by nickel in ambient air based on experimental findings. The CSTEE concluded that the limit value for non-cancer effects should be 20 ng/m^3. This value should also provide "reasonable protection to the

carcinogenic effects of nickel compounds in ambient air for the general population" (CSTEE, 2001).

2.3.3 Summary of literature review

At certain levels, inhalation exposure to arsenic, nickel and possibly cobalt has been associated with increased risks of respiratory cancer. However, with the exception of arsenic, evidence regarding the impact of exposure to the CoCs in community settings is very limited; few studies have examined health effects for types and levels of CoC exposure that exist in Port Colborne.

The evidence linking CoC exposure and risks for cancer and mortality has come largely from occupational cohorts, where the exposures are known to be considerably higher than those from environmental sources. Meanwhile, the environmental exposures in contaminated areas, such as parts of Port Colborne, are higher than those in uncontaminated locations, yet lower than levels observed in most occupational studies. There remains uncertainty as to the possible health effects resulting from exposure to the CoCs at these intermediate levels of exposure. Furthermore, the health effects resulting from exposures to the particular combination of contaminants that exist in Port Colborne are unknown, although a number of studies indicate that they may be different than the effects of exposure to any single CoC. Thus, while the literature supports a potential role of the CoCs in cancer incidence and certain causes of mortality, there is insufficient information to characterize the relationship among individuals as a result of the exposure profile observed in the Port Colborne community.

3. COHORT STUDY DESIGN

A historical cohort study design will be used to address the objectives of the study. The cohort analysis allows for the calculation of cancer incidence and mortality rates within the Port Colborne study population, and for these rates to be compared to those in suitable reference populations. The availability of comprehensive population-based databases permit the research questions to be addressed using existing prospectively collected data. Moreover, these data will allow the health of the residents of Port Colborne to be described from a historical perspective, namely, over the period from 1982 to 2000.

For investigations of cancer in the Port Colborne community, the historical cohort study design is less subject to many of the limitations that may have affected the findings of the previous ecological research (Health Services Department Region of Niagara, 1997; MOE, 2000). The opportunity for potential biases associated with the previous studies is reduced by the use of:

1. Individual-level data, including partial residential histories that are available from population-based registries on an annual basis. With individual-level data, the ecologic bias is avoided, and the residential information can be used to examine

issues of potential bias and misclassification related to mobility of the study population.

2. Information that allows us to incorporate possible occupational exposures received as an employee of INCO (note: it is plausible that other occupational exposures could have been received, but we only will capture those among INCO workers).
3. Various statistical adjustments for other relevant risk factors that may confound the associations. Such risk factors include smoking, occupational exposures and income.

Additionally, the inclusion of 19 years of follow-up data will allow for the identification of a sufficient number of events, thus providing sufficient power to study several health outcomes despite the relative rarity of some conditions. Finally, the proposed study can be performed in a timely manner once record linkage using existing data has been completed.

3.1 Cohort Members

The study cohort will be identified from the T1FF, available for the years 1982 to 2000 (Statistics Canada, 2000). The T1FF was introduced in 1982 for the development and dissemination of social, economic and demographic statistics and indicators for sub-provincial geographic areas (postal areas and selected census areas). The inclusion criteria for the cohort are:

- Adult individuals, defined as equal to or greater than 20 years of age at any point during the follow up. This age cut-point was chosen to correspond to standard age-groupings for Canadian cancer incidence statistics and facilitate comparisons between the cohort and referent populations
- Residency reported in Port Colborne or any one of six selected comparator communities (see below) at any time during 1982 to 2000 inclusive, as determined by a postal code-based definition created for each community. An approximately 50% random sample of individuals from the comparator communities will be chosen.

In addition, since the cohort design requires that follow-up occur among individuals who are free of the condition under study, our analysis of cancer outcomes will exclude individuals who were diagnosed with cancer prior to follow up. Cancer data are available from 1969 onwards and, therefore, our cancer risk assessment will exclude cohort members diagnosed with cancer between 1969 and 1981.

The study cohort will be comprised of individuals who were at least 20 years of age or older at any point in time during the study period. Due to the small population size of Port Colborne, and the rarity of childhood cancer, we will be unable to identify a sufficient number of childhood cancer cases needed to perform statistical analyses. The Canadian age-standardized incidence rate (ASIR) for all child and youth cancers (0 to 19 years of age) is approximately 16.0 per 100,000 individuals (National Cancer Institute of Canada, 2003). The number of individuals aged 0 to 19 years in Port Colborne is estimated from a sample of the T1FF to be 3,356. Applying the above ASIR translates into an expected number of only 0.5 incident cancer cases per year for the study population aged 0 to 19 years.

The period of follow-up will extend from 1982 to 2000. Individuals comprising the cohort must have resided in Port Colborne or any of the other comparator communities for at least one year during the follow-up interval. The residential history of cohort members will be traced as far back as possible. For cancer outcomes, subjects will be followed until the time when first diagnosed with cancer, or the most recent year for which cancer data are available. For mortality outcomes, subjects will be followed until the day of death, or the last year for which mortality data are available. Individuals for whom no record is found in the cancer or mortality files are assumed to be cancer-free and alive, respectively, at the end of the study period.

The viability of using the Canadian mortality and cancer registry database in this manner is well established. Although it is possible a small number of deaths or incidence cancers may be missed, we know of no reason why such cases would be related to exposure status and, therefore, we anticipate no resulting bias in our risk estimates.

The selected comparator communities are several of those defined in the working document, 'Protocol C Technical Methodology Document: The Selection of Ontario Comparator Communities for Port Colborne' (Ventana CRC, 2003b). Discriminant analysis was used to identify a series of Ontario communities that shared similar features to Port Colborne. Communities were identified using census variables that covered a wide range of social and economic determinants of health. Statistical analysis was then performed to identify those communities that were most similar to Port Colborne using 47 census predictor variables. Comparator communities were selected by calculating the Mahalanobis distance (a measure of distance between two points in space that are defined by correlated variables) from each community to Port Colborne. Those communities with the smallest Mahalanobis distances were deemed to be most similar to Port Colborne. Communities with the potential to have elevated CoC exposures or where significant numbers of residents may have worked at an INCO facility were excluded as comparator communities. Based on these criteria, ten communities were excluded from the list of comparator communities (eight communities in close geographical proximity to Port Colborne, the city of Sudbury, and one community in close geographical proximity to Sudbury).

It is important to note that the residential history of individuals using T1FF data is based on postal code information, and not census enumeration areas. Therefore, it is not possible to achieve suitable concordance for some communities identified based on census areas. In consultation with personnel at Statistics Canada, we eliminated those communities where there was poor concordance between population estimates obtained using areas defined by postal codes and census areas. Specifically, those communities where the coverage rates of the T1FF file to the census areas did not fall between 80% and 120% were dropped. As well, data were not available for the entire study period for several of the communities. Six remaining communities were selected where data are available for the entire study period and T1FF coverage fell within the specified range. These communities are listed in

Appendix II, along with the exclusions and an indicator that categorizes two communities with potential environmental exposures relevant to respiratory conditions.

3.2 Cohort Data Source

The development of the T1FF is based on the census family concept used by Statistics Canada. The census family includes parent(s) and non-married children (i.e. who reported a marital status other than 'married' on the tax form) living in the same dwelling. The data contained in the T1FF are based on information contained in annual income tax T1 forms and includes all tax filers as well as any identifiable filing and non-filing family members. Inclusion of non-filing family members in the T1FF file began in 1992. The family formation for the file is done through deterministic and probabilistic matching and imputation. Tax filers from the same family, including children, are first matched using common links (e.g., spousal social insurance number (SIN), same name and same address). The resulting family unit is examined for completeness and if there are indications of non-filing members (e.g., dependent children), those members are added to the file (imputed) based on the information included on the tax forms of the filers in the family. The remaining tax filers who have not been matched in the family formation process become classified as non-family persons. Specifically, non-family persons are those not living with a spouse or child. They may be living alone or with other related family members (e.g., cousin or grandparent) or unrelated persons (e.g., roommate or boarder). Thus, the T1FF contains information for all tax-filing individuals and for most non-filing members of census families containing at least one tax-filer.

The following variables from the T1FF will be used in the analysis: age, sex, birth date, marital status and place of residence as defined by postal code. The T1FF database includes six-digit postal code information. The first three characters of the postal code, known as the Forward Sortation Area (FSA), will be used to identify residential addresses within Port Colborne (the L3K Forward Sortation Area) and the comparator communities. Statistics Canada can also create 'user-defined' areas that are not part of their standard geography level codes. These special user-defined area groupings will be aggregated based on postal codes that correspond to 'special' study areas (e.g., comparator communities not well-defined by an FSA, or areas within Port Colborne corresponding to soil metal concentration contours). Six-digit postal code information will be used to estimate exposure for Port Colborne residents. This will form the basis of internal comparisons of respiratory cancer rates within Port Colborne.

3.3 Cohort Data Quality

Most adult Canadians (approximately 60% between 1978 and 1986, and 70% from 1990 to 2000) file an income tax return in a given year (Statistics Canada, personal communication). In addition, the imputation of non-filing family members increases the coverage of the T1FF. A comparison of the T1FF with census and post-census estimates of population shows that the T1FF only underestimates the Canadian population by

approximately 6% every year, making it an excellent sampling frame for the Canadian adult population (Bleuer, 1996).

Nonetheless, since underestimation by the T1FF is not random, the potential for bias to be introduced by using this source for cohort selection needs to be addressed. In particular, individuals with low or no taxable income are less likely to file taxes and, therefore, will be underrepresented in the T1FF (Statistics Canada, 2000). Non-employed spouses and children form part of this low income non-filing group, but are generally captured by imputation after 1991. Older Canadians receiving only Old Age Security and the Guaranteed Income Supplement are also less likely to file, and since they are less likely to be living with a tax-filing census family member, they are not likely to be captured by imputation. However, the percentage of these individuals filing has increased substantially since 1990, following the introduction of the Federal Sales Tax Credit in 1986, and the Goods and Services Tax Credit in 1989 (Statistics Canada, 2000).

In light of the study outcomes (e.g., cancer) it is important to assess the magnitude of this possible selection bias since, for example, cancer rates are higher among the older age groups. A non-systematic error in selection would only function to reduce sample size (i.e. the power of the study), whereas a systematic error could affect the risk estimate. Since cohort members from comparator communities will also be identified from the T1FF file, underestimation of elderly people and non-filers will also pertain to these comparator communities. Therefore, in the comparison between Port Colborne and the comparator communities, bias would only exist if the coverage of the T1FF file was different across Port Colborne and the comparator communities. It is not expected that differences exist since comparator communities have been selected for their similarity to Port Colborne on a number of demographic factors. External comparisons between Port Colborne and Ontario may be more likely affected by any underreporting bias, given i) the greater differences in various demographic and socioeconomic variables between Port Colborne and Ontario, and ii) Ontario rates have been derived from the entire provincial population rather than linkage of T1FF data to cancer outcomes.

In comparisons of a sample of the T1FF data and census population estimates for Port Colborne, it is noted that coverage of the adult Port Colborne community is high (>95%) for males over the period after 1986. Between 1982 and 1986 for males, coverage is also high (88%) but there is some underrepresentation of age groups greater than 65 years. For females, comparisons to census data indicate an undercoverage of approximately 20% prior to 1992, at which point imputation methods increased female coverage to greater than 90%. However, when examining the overall population for Port Colborne, there are no significant differences in the age-specific population counts of the T1FF sample relative to census data collected from 1986 onwards. Any undercoverage prior to this occurs primarily among the older age groups. These comparisons indicate that the T1FF provide a suitable sampling frame for the adult Port Colborne community.

In comparisons of a sample of the T1FF data and census population estimates for the comparator communities to be used in this study, it is noted that coverage of the

comparator community's combined population is above 100%. The average for the years 1995 to 2000 is approximately 109% (Statistics Canada, personal communication). This is due in part to the delineation of the communities by the two data sources, the T1FF and the census. As noted earlier, communities are defined in the T1FF data by postal code, which does not correspond directly to communities defined by census divisions. The imputation of children may also affect observed discrepancies between the two data sources. Excluding the age group of those aged 0 to 19 years may reduce some of the overcoverage.

3.4 Exposure Assessment

3.4.1 Environmental exposure

An individual's exposure to CoCs is influenced by levels of chemicals in and around his/her residence, activities, behaviour, occupation and food and water consumption patterns. As a result, individual exposure to these CoCs can vary widely throughout the person's lifetime. In our study, there are no direct measures of individual exposure. Our indices will be proxy measures of potential exposure, defined on the basis of residential information (i.e. location, duration of residency and/or soil metal concentration contour level). As such, they will serve only to identify those individuals who **may** be at an increased risk of exposure. This classification is useful, since there is the potential for differential exposure to the CoCs between Port Colborne residents and those of other communities, and among Port Colborne residents.

A number of levels of exposure characterization will be used to address the study objectives and to conduct sub-analyses. A variety of exposure indices can be modelled by constructing a hierarchy of exposure measures. This hierarchical approach begins with a crude classification based on community of residence, and progresses to a more refined classification that incorporates both duration of residence and soil metal concentration. The less detailed measures are based on community of residence, and are designed to create groups that can be compared on the basis of whether or not they lived in Port Colborne. As such, these classifications will permit evaluation of the risks of cancer and mortality experienced by those who ever resided in Port Colborne, directly addressing the primary objective. The more refined exposure classification has been designed such that duration of residency and magnitude of exposure are both incorporated into the risk estimation process. The following section describes each of the exposure measures that will be used. Table 1 briefly outlines the exposure levels, analyses and comparisons. Appendix III contains a more detailed table outlining each exposure measure and identifying for each one its use in the study analyses, and the groups to be compared.

Exposure Level 1: Port Colborne Residence

The primary exposure measure for study analyses will be a time-dependent categorical variable defined by the number of years of residency in Port Colborne for the cohort members between 1982 and 2000. This Level 1 exposure index is assumed to be

representative of a cumulative or lifetime potential for exposure. The residents of Port Colborne have potential environmental exposures to the CoCs that are considerably higher than levels found, on average, throughout the province of Ontario. If a subject's annual residential history between 1982 and 2000 includes, for any year, a Port Colborne postal code, then that individual will be categorized as exposed for that year. Unexposed subjects will be those individuals who did not reside in Port Colborne at any point in time during the study interval. That is, they had accrued zero years of exposure during the study interval.

For each year from 1982 to 2000, each cohort member will be classified as either (1) a Port Colborne resident or (2) not a Port Colborne resident. Their cumulative years resident in Port Colborne will be categorized into groupings representing the total number of years lived in Port Colborne. One categorical representation of this variable could be:

(i) 0 years lived in Port Colborne

(ii) >0 and <5 years lived in Port Colborne

(iii) 5+ years lived in Port Colborne

Other categorical representations can be easily constructed by adjusting the cut points to evaluate difference in risk based on duration of residency. Frequently, the categorization of exposure takes into consideration the number of health events that occur within each level. By so doing, the precision of the risk estimates can be optimized by allowing for sufficient number of cases to be captured within each exposure category.

Table 1: *Hierarchy of exposure assessment*

Level	Exposure Variable[a]		Analysis	Comparison
1	Port Colborne Residence	Categorical variable (number of years lived in Port Colborne). For example, ▪ 0 years ▪ >0 to <5 years ▪ 5+ years	*Primary analysis for comparing:* ▪ Incidence and mortality rates of respiratory cancer ▪ Rates for all cancers and causes of mortality	*External* ▪ Ontario *Internal Cohort* ▪ Comparator communities
2	Average Annual CoC Exposure in Port Colborne	E_{avg} (percentile) ▪ 25th ▪ 75th	*Secondary analysis for exploring:* Associations of CoCs with incorporation of regional variations in potential exposure; Comparing; ▪ Incidence and mortality rates of respiratory cancer ▪ Rates for all cancers and causes of mortality	*External* ▪ Ontario *Internal Cohort* ▪ Comparator communities ▪ Within Port Colborne, Low vs. high[b]
3	Geographic Proximity to INCO Plant	Categories (grouped by quartile) based on residence ▪ lived longest ▪ lived first ▪ lived last	*Secondary analysis for exploring:* Associations between CoCs and incidence rates within Port Colborne	*Internal* ▪ Rates within each quartile

[a] In addition to presenting relative risks for the entire Port Colborne cohort, analyses will also be stratified by a measure of occupational exposure (ever vs. never employee of INCO; see Appendix III).

[b] Sample size permitting.

Although we are lacking residential data prior to 1982, it is not unreasonable to expect that a large portion of our identified 'ever' residents (1982-2000) of Port Colborne lived there prior to 1982 and therefore for an even longer period of time then that assessed over the study period. It has been ascertained from the 1981, 1986, 1991 and 1996 census data that the population of Port Colborne has been quite stable over the decades of the 1970s, 1980s and 1990s (see Table 2).

For each census year, approximately 65% of the Port Colborne population (over five years of age) reported that five years previously they had lived at the same address, and of those who reported moving, more than 20% remained within Port Colborne. This indicates that more than 85% of residents report remaining in Port Colborne after at least five years. From these census estimates, it appears that the majority of individuals identified as living in Port Colborne in a given year are, in fact, likely to have lived there for a longer period, and therefore, will have been potentially exposed to any contamination within the community.

The true cumulative exposure difference between Port Colborne residents identified in the T1FF and the unexposed component of the study population may be larger than is actually being measured. Our exposure measures that only cover the interval between 1982 and 2000 may be a more accurate reflection of exposures over a longer period of time. Extension of follow-up to include years prior to 1982 would be preferred, however, no comprehensive data are available on an annual basis that can be readily linked to the population-based registries.

Although there is variability of potential exposure with place of residence within Port Colborne, this Level 1 exposure measure does not take this variability into account. Rather, it characterizes the risk associated with the variability in potential exposure between the communities (i.e. Port Colborne versus comparator communities) instead of within the Port Colborne community (i.e. across regions of Port Colborne).

Table 2: *Mobility status of Port Colborne residents greater than five years of age*

	Canadian Census years			
	1981	**1986**	**1991**	**1996**
Place of residence five years ago[a]				
a) Total Population	17,925	17,055	17,335	17,230
b) Non-mover (i.e. at same address)	11,820	11,760	11,290	11,990
c) Mover	6,105	5,295	6,040	5,240
d) Non-migrant mover (i.e. moved within Port Colborne)	4,005	3,395	3,835	3,240
Total % in PC five yrs ago = (b+d)/a	**88**	**89**	**87**	**89**

[a] For the five-year mobility question, respondents were asked to write the name of the "municipality and province" of residence five years ago.

Exposure Level 2: Average Annual CoC Exposure

The Level 2 exposure is intended as a secondary measure for additional study analyses. This more detailed exposure level will include surrogate measures of intensity, in addition to the duration of exposure. As noted in Section 2.3, there are different levels of soil CoC contamination within Port Colborne. The variations in soil CoCs are representative of a range of intensity of potential exposures over time. Therefore, soil contour levels mapped on a regional level will be used as a proxy measure of the intensity of potential exposure to CoCs for Port Colborne residents over time. The average annual CoC exposure will be estimated by combining the number of years residing at a particular location (duration) with an estimate of the soil CoC levels at that location (intensity), with the resulting measurement averaged over their residential history in Port Colborne (Figure 1). Nickel contamination provides for the most significant potential community exposure, therefore intensity will be modeled by soil nickel contamination levels and nickel, arsenic, cobalt and copper combined. Location will be represented by the postal code as recorded in the T1FF data.

Figure 1: *Estimate of average annual exposure for an individual while resident in Port Colborne; a composite of duration and intensity (E_{Avg})*

$$E_{Avg}(ppm/yr) = \sum_{i=1}^{k} T_i Y_i \Big/ \sum_{i=1}^{k} T_i$$

Where, i represents each separate residence in Port Colborne (i=1 to k), k represents the total residences in Port Colborne, T_i corresponds to the number of years the subject lived at residence i, and Y_i represents the exposure at residence i (based on soil contour CoC levels).

This exposure variable will be categorized based on the distribution of exposures within the Port Colborne study population. A high exposure category will be defined as exposures falling within the range of the top quartile; a low exposure category will be defined as the lowest quartile of exposure. By defining the exposure groupings in this manner, the potential for exposure misclassification has been reduced. In addition, given the small number of residences located in the highest intensity exposure area, the average annual exposures are expected to follow a skewed distribution, with a large number of very low exposures and relatively few high exposures. By using the highest and lowest quartiles, this should provide two categories with reasonably different exposures, while providing a sufficient number of outcomes to perform the analyses.

Investigations of the health effects of environmental contamination have used environmental monitoring results, such as soil contaminant contour mapping, to identify exposed populations. After investigating environmental methods (e.g., spot sampling and contour mapping) and mathematical methods (e.g., concentric circles) for identifying populations potentially exposed to a point source of airborne pollution, Williams and Ogston (2002), concluded that soil contour mapping should be used to guide the final selection of exposed and non-exposed populations for study. Therefore, including the information available in the soil contour maps produced for each of the CoCs in the Port Colborne area is an important study aspect, as soil contour mapping not only represents current potential exposure from the contaminated soil but also that from historical airborne pollution.

In the review of the literature (see Section 2.3), it was noted that inhalation exposures to nickel and arsenic accounted for the observed increase in risk for respiratory cancer. The follow-up period in this study does not include time during the period of the highest reported levels of fugitive emissions in Port Colborne and, therefore, presumably the time during which inhalation exposures to the CoCs from ambient air pollution would have been greatest. However, due to typical cancer latency periods ranging from 15 to ≥20 years, any cancers resulting from these prior cumulative exposures, including ambient air exposures, would likely be identified during the study follow-up period. As noted above, soil metal concentration contours are a proxy measure of the historical levels of metal

concentrations in the air. In addition, while all measures of residency are based on information available from 1982 through 2000, the assumption is made that for the majority of individuals in the cohort, the more recent records can approximate their residential history within Port Colborne. This study will also include 19 years of follow-up after 1982, therefore an exploration of the risk for respiratory cancer associated with the potential exposure to soil CoCs over this period as represented by current soil contour levels is relevant for current Port Colborne residents and residents of the recent past.

This exposure index will form the basis of three different comparisons. First, Port Colborne residents in the upper quartile of exposure will be compared to the sample of residents in the comparator community. Secondly, rates in the subgroup of highly exposed Port Colborne residents will also be compared to provincial rates. The third comparison will be between those in the upper and lower exposure categories of the Port Colborne cohort.

Exposure Level 3: Geographical Distance from INCO Plant

In epidemiologic studies of environmental pollution, distances from the point source are a frequently used index of exposure. With this design, Port Colborne residential information based on a six-digit postal code on an annual basis between 1982 and 2000 can be used to estimate the distance between each Port Colborne cohort member's residence and the INCO plant. This will be done by taking the geographical distance between the centroid represented by the postal code area and the INCO site. These geographical distances will be categorized according to the observed frequency distribution (i.e. tertiles or quartiles). The number of distance categories will be dependent on the observed number of cancer cases, and the cut-points will be selected to optimize the precision of the risk estimates. Specifically, categorization of distances will ensure there are a nearly equal number of cancer cases within each category. Thereafter, the number of person years spent by the Port Colborne cohort within each distance category will be tabulated. The exposure index will only be used to conduct internal comparisons amongst Port Colborne residents. To examine the possible latency effect for each Port Colborne cohort member, we will estimate risks using geographical distance measures for the residences 1) lived at the longest, 2) first lived at during the follow-up period and 3) lived at most recently. Geographical distances will be estimated using Geographical Information System software, an example of which is Arcview, version 8 (Distributor: RockWare Inc. 2221 East St. #1, Golden, CO 80401).

3.4.2 Occupational exposure

Occupational records will be used to identify cohort members who worked at INCO for at least one year. Statistics Canada will link the study cohort to employment records using unique identifiers included with the INCO data (i.e. name, date of birth and SIN). By utilizing the years individuals' first and last worked at INCO, time-dependent person-years of exposure will be allocated to cohort members using residential information according to employment status. A binary indicator of employment status at INCO (i.e. ever vs. never)

will allow for stratified analyses to be conducted. This will allow us to separate out occupational and environmental exposure effects and, additionally, to examine the combined influence of occupational and environmental exposures. Since there are no comparator communities in the vicinity of Port Colborne, and Sudbury has been excluded as a comparator community, it is unlikely that anyone from the unexposed group will have employment status at INCO. Nonetheless, linkage of individuals from the comparator communities to the INCO database will identify any such individuals. Due to occupational transfers within companies, Port Colborne INCO employees may have been Sudbury INCO employees at one time and to the extent that this is the case, such Port Colborne residents may be at an increased potential for exposure to the CoCs from their Sudbury resident status. With the availability of the years first and last employed at INCO, such individuals would be readily identified. The potential for possible exposures in Sudbury will be characterized for the Port Colborne cohort members and taken into consideration in sensitivity and sub-analyses.

The stratified analysis of the cohort will provide an estimate of risk in Port Colborne residents excluding those identified as employees of INCO for at least one year. In this way, it is presumed that factors contributing to the observed rates and estimated relative risks will exclude those individuals with potentially high occupational exposures. INCO employees may have been more likely to reside in the East Side of Port Colborne, in proximity to the INCO plant where CoC levels are much higher. Therefore, for this region, we would have data for individuals exposed to high levels of occupational and environment exposure, and for individuals (e.g., spouses) with no occupational exposure and high environmental exposure. This variability in exposure will permit us to examine the separate roles of environmental and residential exposure.

Analyses of the Level 1 Port Colborne residence exposures will be stratified on INCO employment as stated above. In addition, a slightly more complex variable containing both occupational and residential information will be constructed and used as part of the Level 2 and Level 3 sub-analyses within the Port Colborne cohort. Four exposure categories will be created, based on a dichotomous occupational measure and a dichotomous residence exposure variable. Table 3 describes possible exposure categories.

Table 3: *Cohort exposures including occupational component*

Exposure Category	Long-term PC Residency (Level 2)		Average Exposure (Level 3)	
	Residency [a]	Occupation [a]	Residency	Occupation [a]
1[b]	Low	Low	25th percentile	Low
2	Low	High	25th percentile	High
3	High	Low	75th percentile	Low
4	High	High	75th percentile	High

[a] Low is < five years and high is ≥ five years.

[b] Group 1 represents the referent category.

3.5 Outcomes of Interest

3.5.1 Cancer registry data

Through provincial cancer registries, Canada is one of the few countries in the world that has a cancer reporting system that covers the entire population. Through the National Cancer Incidence Reporting System (NCIRS), the Canadian Cancer Registry (CCR) has compiled Canadian cancer incidence rates dating back to 1969 (Band et al., 1993; Gaudette and Lee, 1997). The CCR is a longitudinal, person-oriented database containing information on all Canadian residents (permanent and non-permanent) diagnosed with cancer (LaBillois, 1999). Data for the years up to 2000 will be available for this study (Statistics Canada, 2003).

Provincial and Territorial Cancer Registries (PTCRs) are principally responsible for the degree of coverage and the quality of the data (LaBillois, 1999). Canada-wide, more than 82% of cases have been microscopically confirmed since 1969. In Ontario, only 74 % were microscopically confirmed between 1969 and 1973, but by 1984 that proportion had risen to 88% (Band et al., 1993). The rate of microscopic definition for respiratory cancer is quite good at 78%. For other cancers where differentiation is important (e.g., non-Hodgkin's lymphoma versus leukemia) histologic confirmation is more critical. Data on newly diagnosed cancer cases is provided on an annual basis and files include the following types of information (Gaudette and Lee, 1997):

- Patient name
- Health insurance number
- Provincial registry identifier
- Place of residence name (standard geographic code of residence)
- Birth place
- Sex
- Patient status
- Date of birth
- Method of diagnosis (e.g., microscopic, death certificate only, etc)
- International Classification of Disease, Tenth Revision (ICD-O) topography and morphology
- International Classification of Disease, Ninth Revision (ICD-9) code of diagnosis
- Primary site number
- Date of diagnosis
- Death registration number (where relevant)
- Date of death (where relevant).

All incident cancers (i.e. primary malignant neoplasms) within the study cohort will be identified by linking personal identifying information from the T1FF file to the CCR over the period from 1982 to 2000. The cohort will also be linked to the CCR to identify those

individuals with a diagnosis of cancer that occurred before the start of follow-up (between 1969 and 1981). These individuals will be excluded from the cancer incidence follow-up. Our analysis will only consider the first diagnosis of a primary malignancy.

Respiratory cancers for the primary study objective will be defined as primary malignant neoplasms of the: (i) nasal cavities (ICD-9 160), (ii) trachea, bronchus and lung (ICD-9 162) and (iii) larynx (ICD-9 161). Risk estimates will be calculated for all respiratory cancers combined (ICD-9 codes 160-162). It is customary for cancer registry reports to exclude non-melanoma skin cancers, and benign tumours (National Cancer Institute of Canada, 2003). Therefore, due to the potential for incomplete data on these outcomes, and to facilitate appropriate comparisons, the analysis for all cancers combined will exclude non-melanoma skin cancer (ICD-9 173). In addition, benign neoplasms will not be included. Benign tumours (apart from those of the brain) are typically excluded from epidemiologic studies as they are frequently asymptomatic, their development is unrelated to the process of carcinogenesis and their health consequences are generally not severe.

3.5.2 Vital Statistics Death Database

Death registration in Canada is the responsibility of the provinces and territories, under a federal-provincial-territorial agreement. The national Vital Statistics Death Database collects data annually from all Canadian provincial and territorial vital statistics registries. The database contains all deaths that occur in Canada, as well as the deaths of Canadian residents that occur in some American states. Data for the years up to the end of 2000 will be available for study (Statistics Canada, 2003).

In Canada, the ICD-9 was used to code the underlying cause of death between 1979 and 2000. Linkage to the Vital Statistics Death Database will provide the date and underlying cause of death for deceased cohort members. Those persons for whom a death record is not found will be assumed to be alive at the end of the follow-up date (December 31, 2000). For 2000, underlying cause of death data will be recoded from ICD-10 to ICD-9 to maintain consistency with identified deaths in the early part of the study interval.

4. STUDY METHODOLOGY

4.1 Database Linkage

4.1.1 Study outcomes

The study cohort will be linked to mortality and cancer incidence data by using a probabilistic procedure referred to as the Generalized Record Linkage System (Statistics Canada, 1993). Personnel at Statistics Canada will perform the record linkage. The Generalized Record Linkage System compares common fields in the two files to be linked, assigns weights to the resulting links and calculates a total weight. Links with a sufficiently high weight are accepted as a match. Records from the T1FF and the cancer and mortality databases will be linked using personal identifying information (i.e. name,

sex, date of birth and SIN). The presence of the SIN data in the T1FF database will allow for an internal linkage within that database that captures individuals who have changed their name during the follow up period (e.g., after marriage). The SIN information would be used only for this linkage, as it is not present in the cancer or mortality databases. Based on the availability of unique common identifiers available in the data sources for study, the Statistics Canada record linkage between files is expected to have a high level of accuracy.

Data from the CCR have been used extensively in record-linkage studies (Howe and Lindsay, 1981; Terry et al., 2002; Band et al., 2001; Jain et al., 2000a; Jain et al., 2000b; Hertzman et al., 1997; Finkelstein, 1996). A high level of confidence can be placed on the validity of each cancer case identified, since approximately 85% of cancer diagnoses within the CCR are confirmed by microscopic examination of tissue via autopsy, histology or cytology (Band et al., 1993). The proportion of diagnoses that are microscopically confirmed does vary by cancer site with the less accessible sites, and those less likely to be biopsied for other reasons, having lower proportions. Among the respiratory cancers that are of primary interest in this study, lung cancers have a slightly lower proportion of microscopic confirmations (from 72% in 1969 through 1973 rising to 78% after 1984), while the proportion of esophageal cancers that are microscopically confirmed is similar to the overall numbers. Cancer of the larynx has a high proportion of microscopic confirmation (90% in 1969 through 1973 and 97% after 1984).

Due to the completeness of the CCR data, the linkage to the CCR will capture most, if not all, first diagnoses of cancer that have occurred in cohort members. The linkage will even capture diagnoses of cancer that have occurred in cohort members within Canada, but outside the province of Ontario. While this is not expected to be a large number, the ability to ascertain these cases will minimize the possibility of bias associated with incomplete ascertainment of cancer cases. Similarly, linking to the death records for analyses is expected to have a high degree of completeness, capturing over 95% of deaths that have occurred in cohort members. Any incomplete ascertainment of cancer incidence or mortality is not expected to differ by exposure status. Our risk estimates should not, therefore, be impacted by any missed cases. Note that the inability to track all cases will not change the estimate of risk assuming loss to follow-up is non-differential with respect to exposure. However, missing cases will affect the precision of our risk estimates (i.e., 95% confidence intervals, and in this respect, affects our results).

4.2 Study Comparisons

The primary study objective will be investigated by comparisons of respiratory cancer incidence and mortality. Our definition of respiratory cancer includes malignancies of the lung, nasal sinus passage, trachea, bronchus and larynx. Secondary objectives will be addressed by exploring external comparisons of all cancers and all causes of death, and internal comparisons of respiratory cancer incidence rates. External comparisons will

include Port Colborne residents relative to a group of comparator communities selected from Ontario and to population-based rates calculated for all Ontario residents.

4.2.1 External comparison

The Ontario population provides a large sample size for comparing cancer incidence and mortality rates calculated for the Port Colborne cohort. However, given that a large proportion of the total population of Ontario resides in metropolitan areas, Ontario data are highly influenced by the rates observed for these areas. Therefore, an external comparison to a group of comparator communities will be performed in evaluating the primary study objective.

External comparisons to Ontario will be conducted by multiplying the tabulated number of age-sex specific person-years for the Port Colborne cohort by published age-specific cancer rates for Ontario. This allows us to calculate the number of cases of incident cancer that would be expected if Port Colborne residents had the same rates as the Ontario general population. Through record linkage, we are able to identify the observed number of cases in the Port Colborne cohort. By dividing the observed number of cancer cases by the expected number, we are able to calculate the SIR. Using the same methods, we can also calculate the SMR. Tests of significance will be constructed by evaluating the 95% CI of the SIR or SMR.

4.2.2 Internal comparisons

In addition to the external comparison, variations in respiratory cancer incidence and mortality rates will be explored within the study cohort (Port Colborne versus comparator communities) and geographically within Port Colborne. Statistics Canada data demonstrate considerable variability in sociodemographic characteristics of many communities within Ontario (Statistics Canada, 2002). Many of the community characteristics (e.g., income level, education, employment status and ethnicity) are related to cancer incidence and mortality. Therefore, the comparison of Port Colborne rates to sociodemographically-matched communities allows many of these differences to be taken into account. By matching selected demographic characteristics of the Port Colborne community to a sample of other Ontario communities, we are able to adjust for the influence of these factors in the comparison of cancer incidence and mortality rates between the two populations. If differences in cancer incidence and mortality rates remain after controlling for these factors, then observed differences may be attributed to factor(s) that are unrelated to the sociodemographic profile of the community (e.g., environmental exposure) with greater confidence.

To allow for an examination of where those diagnosed with cancer lived within the community, the location of residence for each cancer case will be mapped according to: (1) last address in Port Colborne prior to time of diagnosis; (2) address at which residence duration was longest (prior to diagnosis); and (3) address first lived at during the follow up period. This is an important aspect of the study since there has been potential differential exposure to contaminants within the community.

Sample size permitting, regional comparisons will be made as outlined in Appendix III.

4.3 Potential Confounders

This study has been designed to examine the effects of residential exposure to the CoCs on a number of mortality and cancer outcomes. However, the main focus of the study is on respiratory cancers, as these have been previously linked with exposure to one or more of the Port Colborne CoCs. Therefore, any strong risk factors for respiratory cancers should be examined as potential confounders. In addition, other known risk factors for respiratory cancer or strong risk factors for other outcomes may be examined as potential confounders.

4.3.1 Demographic variables

Age and sex are strongly associated with many diseases and outcomes, including most types of cancer and causes of death. They are also associated with many other risk factors, including behavioural and environmental factors and, therefore, are commonly found to act as confounders.

Income and other measures of socioeconomic status (SES) have also been found to be predictors of many health-related outcomes, and are clearly associated with many environmental, behavioural and social factors. Income level will likely be of particular interest for this study, as it is likely to be related to residential history.

Information will be available within the study cohort on an individual level for the variables of age, sex and marital status (another measure of SES). These four variables will be examined for their role as potential confounders for all disease outcomes. Summary measures of mean income, by six-digit postal code, will be estimated and considered as a potential confounding variable. This income index is correlated with many services delivered at a community or neighbourhood level (e.g., education and health care) that are important determinant of health. Where appropriate, this information will be used in multivariate modeling to control for confounding. We will also explore the potential for effect modification by age and sex.

4.3.2 Smoking

Smoking has the potential to be a confounder of many disease-exposure relationships, including those to be examined in this study. There is a very strong association between smoking and respiratory cancers, especially lung cancer, and also between smoking and cardiovascular disease (USDHHS, 1989). Smoking is also associated with various respiratory diseases and with a number of non-respiratory cancers (USDHHS, 1989). Therefore, differences in smoking rates between populations can result in observed differences in disease risk between those same populations.

Smoking is a well-recognized correlate of SES, which implies it will tend to be related to other relevant risk factors (e.g., diet and occupational exposures). In the current context, it is plausible that smoking may be related to exposure to the CoCs through the relationship

between SES and residential location. As such, it has the potential to be acting as a confounder of the CoC-disease relationship. It is also important to note that confounding by smoking could mask or create the appearance of either excess or a deficit of a disease outcome, depending on the relationships between smoking, disease and exposure to CoCs. The results of this study will therefore be useful only if there is some way of assessing the degree of confounding by smoking and of controlling for the effects of any such confounding.

Our proposed approach is to use several different strategies to evaluate the possible impact that smoking may have on our observed risk estimates and, where appropriate, to adjust the risk estimates accordingly or to report the degree of bias that may be involved. Ideally, we would control directly for smoking by obtaining detailed individual level information on smoking history and using this information to adjust the statistical models of exposure-disease relationships. However, there is no data source with such information that could be linked to the study cohort, and contacting cohort members to inquire about smoking history is not possible due to both privacy restrictions and the fact that many of those who were diagnosed with cancer have since died. As an alternative, we will use six indirect methods to evaluate the role of smoking on our risk estimates as described in more detail below. In brief, these methods include:

1. Making adjustments in the regression analyses for smoking at a group level using health survey data.
2. Examining rate ratios for various disease outcomes (e.g., smoking and non-smoking related health conditions) between our comparison groups.
3. Adjusting our risk estimates for other recognized correlates of smoking.
4. Conducting sensitivity analyses to evaluate the possible bias in our risk estimate based on plausible differences in smoking according to exposure status (Greenland, 1996; Thomas, 1987).
5. Performing a stratified analysis by gender.
6. Performing analysis by regressing community-specific respiratory cancer disease rates against a community-specific summary measure of smoking prevalence. The predicted respiratory cancer rate for Port Colborne using this linear model would then be compared against its observed cancer rate.

Adjustment for Smoking at a Group Level

Smoking rates are available from the Ontario Health Survey (1990 and 1996-97) at the public health district level, by age/sex group. Hierarchical regression or Generalized Estimating Equations (GEE) methods will take into account the potential confounding influence of ecologic rates of smoking using the available survey data.

Smoking-Related Causes of Death

Smoking is the leading cause of premature mortality in North America (Collishaw et al., 1988; Ellison et al., 2000; USDHHS, 1990), and has been shown to be a major risk factor for several cancer sites (lung, larynx and oral cavity, esophagus, bladder, kidney and

pancreas), and causes of death (cardiovascular, respiratory) (USDHHS, 1982; USDHHS, 1983). The degree to which smoking increases the risk for disease has been estimated in prospective studies of smokers for a number of these conditions (USDHHS, 1990; Steenland et al., 1984). Elevated rates of smoking among a cohort may produce elevated levels of these diseases, as compared to a reference population. In such a case, excess risks would be expected for each of these disease outcomes. If the excess risk is observed for only one of the outcomes, this may be taken as evidence that an excess of smoking among the cohort members is not the explanation for the observed increase in the specific disease risk (Steenland et al., 1984). Furthermore, in the case where elevated risks are seen for a number of causes of death, the relative elevation of each of the smoking-related causes of death can be evaluated, and can provide evidence as to whether or not smoking rates in the population appear to be affecting mortality rates.

Controlling for Other Variables Related to Smoking

Income is related to smoking, and average income levels for each six-digit postal code data will be tabulated and applied to the residential histories of each cohort member. Adjusting for individual level income will not be done as privacy restrictions preclude the release of such data from the T1FF file. Furthermore, the effect of adjusting for this income index can be examined, compared with the estimated associations between income and smoking, and used to evaluate to what degree controlling for income is also controlling for smoking. Comparison of rates within Port Colborne to those in the comparator communities will also provide the same sort of indirect control for smoking, as by design, these communities were selected according to several socioeconomic variables that may be related to smoking patterns within a community.

Sensitivity Analyses

Microsimulation techniques, a type of sensitivity analysis, can be used to estimate the effect of differential smoking rates between two populations on the risk estimates. This is achieved by simulating a large number of cohort study populations. Smoking characteristics for the individuals within the cohort would also be simulated by defining the underlying distribution for a smoking variable of interest (e.g., smoking status). This distribution can incorporate differences in smoking characteristics between communities, and postulated relationships between exposure and disease (i.e. respiratory cancer). By simulating several different scenarios, an estimate of the possible magnitude of bias can be made.

Evaluating Effects Through Stratified Analyses

Stratified analyses by gender will be used to examine relative risks for cancer in males and females. If the role of behavioural (i.e. smoking) and occupational factors is of little consequence, then any observed increased relative risks that are similar between males and females may be attributed to a common shared environmental exposure.

Regression Analysis of Community-Specific Respiratory Cancer Incidence Rates

Linear regression analysis will be conducted by creating an age-sex summary smoking measure for each of the comparator communities and Port Colborne as an independent variable. An age-sex standardized measure of respiratory cancer incidence rate will be the dependent variable. Using these 34 observations, the model will be used to predict the respiratory cancer incidence rate for Port Colborne. This will be compared to the observed respiratory cancer incidence rate based on the identified cancer cases and person-years of follow up within the Port Colborne cohort. Excess in this observed rate would support the hypothesis that factors other than smoking (e.g., environmental) are associated with an increased incidence of this malignancy.

5. SAMPLE SIZE AND STUDY POWER

The number of health outcomes observed during follow-up is the primary determinant of study power in a cohort study design. For cancer studies, the number of observed cases will be influenced by the age and sex distribution of the population at risk. Naturally, with a longer follow-up there exists a greater opportunity to identify a larger number of incident cancer cases. The power of such a cohort study will be influenced by changes in disease or death rates that occur during the follow-up interval. In our study, respiratory cancers represent the outcome of primary interest, and therefore, form the basis of our power calculations.

In order to estimate study power, we must calculate the expected number of outcomes. For this study cohort, this would involve applying age-sex and site-specific rates of respiratory cancer to the expected number of individuals at risk as they are followed up over time. Because this study is a historical cohort design, a more precise estimate of study power can be calculated by applying such rates to the cohort on an annual basis.

The primary objective of this study is to compare the patterns of incidence of respiratory cancer among Port Colborne residents with a minimum residency time of one year (as recorded in the T1FF) to those of a comparative population. Two comparative populations will be used: a sample from several comparator communities from Ontario and the general population of Ontario as a whole. Using Ontario cancer incidence rates, the estimated numbers of site-specific cancer cases that would be identified among residents of Port Colborne are presented in Table 4. It is important to note that there are subtle differences in the power calculations for the two comparisons. For the comparator communities, the estimates risks will be derived from a subgroup of the cohort. In contrast, the comparison to Ontario will make use of external rates published from population-based registries. A succinct account of the differences with respect to sample size when comparisons are performed using external (e.g., Ontario) and internal (e.g comparator communities) comparison groups is found in Breslow and Day (1987).

Table 4: *Estimated number of incident cancers among Port Colborne residents, at time of diagnosis, by cancer site, 1975 to 1999*

Cancer Site	Estimated Total Incident Cases (1975 to 1987)	Estimated Total Incident Cases (1988 to 1999)	Estimated incident cases (annually)
All cancers	1258	1051	92.36
Trachea, bronchus and lung	158	189	13.88
Colorectal	158	174	13.28
Female breast	133	166	11.96
Prostate	86	168	10.16
Bladder	53	48	4.04
Leukemia	34	36	2.80
Pancreas	29	30	2.36

Source: Unpublished tabulations, Health Canada, 2003

5.1 Comparison to Rates of Selected Comparator Communities

There are approximately 14,000 adults residing in Port Colborne, while annually there are approximately 14 incident cases of respiratory cancer. Therefore, the crude incident rate for respiratory cancer is near 1/1,000 in Port Colborne and is similar to Ontario (Health Canada, personal communication, August, 2003). Assuming that the comparator communities experience this same rate, we can readily calculate the power to compare rates between Port Colborne and the comparator communities by estimating the number of outcomes in both populations. In section 5.2, we estimate the number of incident cancers in Port Colborne. Given the disease rate and the estimated cumulative number of cancer cases (n=248), we can estimate the number of person-years of follow-up during the 19-year study period as follows:

$$PY = 248/0.001 = 248{,}000$$

Therefore, for each Port Colborne cohort member we estimate there are, on average, approximately 17.7 (248,000/14,000) years of follow-up. Assuming that the average length of follow-up time is equivalent for sampled residents in the comparator communities, we can also estimate the number of person-years at risk for cohort members that resided in the comparator communities. Based on coverage of the T1FF file, with a 50% sampling, there are an estimated 123,900 adults in the six comparator communities. This comprises 2,193,030 person-years of follow-up (123,900 × 17.7). By assuming that the outcomes are Poisson-distributed we can approximate the power of the study using the normal distribution and the fact that the variance of a Poisson-distributed variable is equal to its mean. In doing so, the comparison of respiratory cancer incidence rates among Port Colborne residents to those who resided in the comparator communities would have a power of 80% at an alpha of 5% to detect a rate ratio of 1.21. Even if we factored in a

liberal adjustment for missing residential data in the T1FF (50%), the study would still have a power of 80% at an alpha of 5% to detect a rate ratio of 1.30.

5.2 Comparison to Ontario Rates

For an external comparison to Ontario rates, the SIR will be the statistic used to compare cancer incidence rates among Port Colborne residents. This measure of relative risk is defined as the observed number of incidence cases (I), divided by the expected number of cases (E). Tests of significance of the SIR or confidence limits can be used to evaluate whether Port Colborne has different cancer incidence rates than the comparators. It is assumed that the number of incident cases of cancer follows a Poisson distribution. Similarly, the SMR will be the statistic used to compare disease-specific death rates between populations.

To calculate study power for the primary study objective it is necessary to calculate the expected number of respiratory cancer cases among 'ever' residents of Port Colborne. The T1FF data covers approximately 94% of the Port Colborne population, therefore we estimate that the total number of cancers of the trachea, bronchus and lung is 248 (19 years of follow-up × 13.88 cases per year × 0.94 coverage = 247.9). We applied the sample size formulae for comparing rates to an external standard (here Ontario) as outlined by Breslow and Day (1987). It is important to note that these power calculations differ from those of the comparator communities as the latter represent a control group within the study cohort. Sample size calculations for the Ontario comparison are based on comparing the observed number of cases in the Port Colborne part of the cohort to the expected number had these cohort members experienced the same rates as the province of Ontario. With an expected number of 248 respiratory cancer cases, we will have 80% power to detect a rate ratio of 1.12 at an alpha of 0.05. Therefore, the study will have ample power to detect a meaningful level of estimated risk in relation to Ontario rates.

The study sample size for the sub-analysis that imposes the five-year residency requirement is also expected to have sufficient power to make comparisons to Ontario rates. For example, the estimated total number of cancers of the trachea, bronchus and lung is one half of the number of incident cases calculated above for a cohort defined by a five-year cumulative residency requirement (assuming 50% of Port Colborne residents live in the area for at least five years). Therefore, with an expected number of 124 respiratory cancer cases, the finding of an observed number of cases in excess of 146 will be statistically significant with 80% power for an alpha of 5%. This translates into a rate ratio of 1.17 (Table 5).

Additionally, a similar calculation can be performed to estimate the number of expected incident respiratory cancer cases during the follow-up period of a cohort defined by the average exposure estimate based on duration and place of residency. If the 75th percentile of the average index defines the grouping for comparisons, then the expected number of incident respiratory cancer cases will be one quarter of that expected for the entire cohort (i.e. defined by a minimum one year residency in Port Colborne). Therefore, with an

expected number of 62 (248/4) respiratory cancer cases, the minimal detectable risk ratio that could be detected with a power of 80% and alpha of 5% is 1.24 (Table 5).

Table 5: *Minimal detectable relative risk based on a study power of 80% and a two-tailed alpha of 5% for Poisson distributions with selected mean values (for external comparison to Ontario standard rates)*

Estimated number of cases	Cases based on power of 80% and Type I error of 5%	Minimal relative risk (two-tailed)
248	278	1.12
124	145	1.17
100	119	1.19
62	77	1.24
50	63	1.26

*based on comparison to an external standard using formulae by Breslow and Day

5.3 Comparison Amongst Port Colborne Residents

A secondary objective of the study is to compare rates of respiratory cancer across residents of Port Colborne with differences in imputed levels of exposure. If we compared the upper 25% percentile of adults with exposure to CoCs (25% of 14,000 = 3,500) to those with the lowest 25th percentile (n=3,500), we could calculate the minimally detectable relative risk as above. Specifically, there would be 61,950 (3,500 × 17.7) person-years of follow-up in each of these two quartile groups. This comparison of respiratory cancer incidence rates among Port Colborne residents in the highest quartile of exposure, relative to the lowest, would have a power of 80% at an alpha of 5% to detect a rate ratio of 1.52 assuming the incidence rate was 0.001. However, if there are only 10 years of follow-up available for each individual, the minimally detectable risk increases to 1.70; this increases to 2.02 if there are only five years of follow-up available on average. Therefore, the study has much weaker power to evaluate any difference across Port Colborne residents according to exposure status.

5.4 Additional Comments on Power Calculations

The calculation of study power involves making several assumptions about disease rates and the average length of follow-up for cohort members. These assumptions are made in the absence of data that could only be obtained after the record linkage has been conducted.

To the extent that incidence rates of respiratory cancer are higher in Port Colborne and the other comparator communities relative to Ontario, our power estimates will be understated. This may be the case given that these are industrial communities, possibly with higher rates of smoking relative to Ontario.

Our power estimates do not take into account the effects of confounding factors such as age, sex and smoking status. For this reason, they may overestimate the ability of our risk estimate to detect differences after adjusting for these variables.

We have not estimated the study power for mortality outcomes. The disease of primary interest, respiratory cancer, is associated with a poor prognosis as less than 10% of individuals diagnosed with this malignancy will survive three years after diagnosis. Therefore, the number of mortality events will be slightly less than the corresponding incident events. As a result, the power to detect difference using mortality outcomes will not be severely comprised.

Finally, it is important to note that we have likely underestimated the number of cases of incident cancers in Port Colborne presented in Tables 4 and 5 for two important reasons. First, the estimates are based on the place of residence at the time of diagnosis. Our cohort will include identified cases from Port Colborne that were diagnosed after they had moved out of the area. Second, this estimate only includes respiratory cancers of the trachea, bronchus and lung (ICD 162). To the extent that other respiratory cancers are included, we will have even greater study power. However, such an improvement would be quite modest as cancers of the trachea, bronchus and lung account for the vast majority of respiratory cancers.

6. METHODS OF ANALYSIS

6.1 Data Analysis Strategy

SIR, SMR and Poisson regression methods will be used in the analysis of the data. SIR and SMR will be used to compare respiratory cancer incidence, all cancers combined and mortality from selected causes of death among Port Colborne residents and a sample of residents from a series of comparator communities. The incidence of respiratory and other forms of cancer will be studied in relation to potential exposure to the CoCs in Port Colborne, where sample size permits. The observed number of health outcomes will be compared to the expected number calculated under the assumption that the Port Colborne cohort experienced rates observed in the province of Ontario. For Ontario comparisons, significance testing of the observed SIRs and SMRs will be done by Chi-square statistic and 95% confidence intervals.

For internal cohort comparisons, Poisson regression modelling will be used to determine the risk of occurrence of cancer or death at a given time in relation to the potential for exposure to CoCs as defined by residence in Port Colborne. These internal cohort comparisons will involve comparing rates across Port Colborne and comparing rates in Port Colborne residents to those of the comparator communities. Regression models will be controlled for a limited number of confounding variables (e.g., marital status, income). The accuracy of the standard errors of the rate ratios derived using Poisson regression will be evaluated by examining whether there is overdispersion in the data. If this is the case,

the standard errors of the risk estimates will be corrected using either GEEs, a scaled deviance or by fitting a negative binomial model. The negative binomial model is an extension of the Poisson regression model and includes an additional parameter that can directly estimate dispersion in the data.

6.2 Characterization of Study Cohort

The age, sex and household size characteristics of the cohort, by community, will be compared to the corresponding Canadian census data. For each individual under study, we will have annual residential information available between 1982 and 2000. For Port Colborne and each comparator community, these characteristics will be compared to 1986, 1991, 1996 and 2001 census data. This will enable us to evaluate the completeness of the T1FF file and characterize the coverage for Port Colborne and the comparator communities over time. The residential mobility of the T1FF cohort members, both for Port Colborne and other residents, will be described. Specifically, the frequency distribution of the number of moves during the 19 years of follow-up (0, 1, 2, 3, or 4+) will be tabulated. We will also identify the proportions of residents who lived in Port Colborne and the comparator community group for at least three, five, seven and ten years.

6.3 Estimation of Person-Years of Follow-Up

Due to the longitudinal nature of the study, we will calculate the number of person-years of follow-up for each member of the T1FF cohort. Person-years will be tabulated across categories of age, sex and community according to annual residential histories as defined by postal codes in the T1FF file. This will also permit a cumulative index of exposure to be created. This index of exposure would be a time-dependent covariate that represents the total number of years that each cohort member lived in Port Colborne. It is termed time-dependent as it takes into account changes in the value of this variable as each individual is followed up over the 19-year study period. The person-years of follow-up will be stratified by age-group (<20, 20 to <25, 25 to <30, 30 to <35, …, 75 to <80, 80 to <85, and 85+), sex, family income level, year (1982, 1983, …, 2000), community and total number of years resident in Port Colborne. Age group and income level also represent time-dependant variables.

By tabulating the number of years of residency in Port Colborne, we have tremendous flexibility in the creation of several residency-based exposure indices. For example, we could define exposure across three categories:

1. No Port Colborne residency
2. >0 but <5 years residency in Port Colborne
3. At least five years of residency in Port Colborne

We are able to construct such time-dependent exposure indices due to the availability of postal code information on an annual basis within the T1FF file. For those individuals who are missing data for a given year, this person-year will be allocated to an unknown

residency code. The above three level exposure variable can be modified to examine the impact of different residency restrictions on our risk estimates. We will conduct such sensitivity analyses by considering three and 10 years of residency as cut points in addition to the five-year restriction cited above.

By using the time-dependent residency variable denoted by the six-digit postal code and their associated measures of CoCs, we will also create an average exposure index based on person-years of follow-up that incorporates intensity of CoC exposure based on soil samplings. Additionally, as discussed previously, we will tabulate the number of person-years based on categories of geographical proximity to the INCO plant. These categories will be defined according to observed distribution of distances among the Port Colborne portion of the cohort. The cut-points of this categorization will be done such that there are a sufficient number of cases to derive stable risk estimates. All of these approaches to assigning exposure using person-years are traditional methods used in occupational epidemiology (Rothman and Greenland, 1998; Monson, 1980; Breslow and Day, 1987).

It is possible that some cohort members will have lived in both Port Colborne and one (or more) of the comparator communities during follow-up. However, we expect that the percentage of such cohort members to be quite small, particularly since the selection of comparator communities excluded those that were in close geographic proximity to Port Colborne. Nonetheless, the use of a time-dependent covariate as discussed above will allow for the exposure profile of such individuals to be appropriately captured. Statistics Canada will review the data file to ensure that each observation represents a uniquely defined individual.

The follow-up will start at the earliest year that an individual is identified in the T1FF file. For the comparison of cancer rates, follow-up will extend from the date of entry into the study cohort until the earliest of date of cancer diagnosis or the last day of the last year for which cancer data are available. The date of entry into the cohort will be the first year the individual was at least 20 years old and lived in Port Colborne or one of the comparator communities. Similarly, for the comparison of mortality outcomes, follow-up will extend until the date of death, or to the last day for which death data are available. Person-years will be allocated to the appropriate age grouping by taking into account changes in age during follow-up. Person-years will be tabulated using the DATAB module of the software program Epicure (HiroSoft Corporation, Seattle, Washington).

6.4 Comparison of Rates between Port Colborne and the Comparator Communities

Poisson regression analysis will be used to compare cancer incidence and mortality rates between Port Colborne cohort members and the sample of residents of six comparator communities. The dependent variable will be the observed counts of cancer incidence or mortality, while the offset or rate multiplier will be the number of person-years. As before, the data will be partitioned according to age, sex and calendar year. Exposure will be treated as residency in Port Colborne and, therefore, coded as a dichotomous variable

(0=comparator community, 1=Port Colborne). We will conduct stratified analyses of cohort members on the basis of INCO employment status (as outlined in section 3.4.2).

We will estimate the rate ratios and their accompanying 95% CIs to assess whether differences in disease rates are statistically significant. Again, these rate ratios will be calculated for each health outcome under study, and also for the different residency restrictions. The SAS procedure GENMOD (SAS, 2002) will be used to estimate the rate ratios. If the count data are correlated (overdispersion is present) statistical methods will be applied to adjust the standard errors of the risk estimates. Specifically, GEEs, scaled deviance or the negative binomial model will be applied to evaluate the accuracy of the confidence intervals produced using the Poisson models.

6.5 Comparison of Rates among Port Colborne Residents to Ontario

In order to compare the cancer and mortality patterns of Port Colborne to Ontario, suitable referent data are needed. Age-sex specific cancer incidence rates will be obtained for the general Ontario population for each year between 1982 and 2000. These rates will be obtained for each cancer site under study as well as for all cancers combined. These rates will be calculated using data provided by the provincial cancer registry and population estimates from Statistics Canada. Age categories will be defined according to five-year groupings.

Similarly, age-sex and period-specific rates for the Ontario population will be calculated for each cause of death under study. A list of causes of death to be examined and the corresponding ICD-9 coding is included in Appendix IV. As before, population estimates will be obtained from Statistics Canada's census and intercensal figures. The observed number of deaths in our study cohort will be obtained by linking our cohort to the mortality database of Statistics Canada.

The number of person-years among Port Colborne residents which have been cross classified by age, sex and calendar year will be multiplied by the corresponding population-based cancer incidence and mortality rates for Ontario. This calculation will yield the expected number of outcomes under the assumption that these residents experience the same rates as the Ontario population. Through record linkage of the T1FF Port Colborne cohort, we will be able to identify the observed number of such outcomes. We will then calculate the SIR for cancer outcomes, and the SMR for mortality outcomes. The 95% CI for these ratios will be calculated by assuming that the occurrence of these health outcomes follows a Poisson distribution. These confidence intervals will enable us to determine whether or not there are statistically significant differences in rates between the two populations. These SIRs and SMRs will be calculated for each cancer and mortality outcome under study. They will also be calculated using person-years and count data for the following minimum residency requirements: no restriction, three years, five years and ten years. Because of the time-dependent nature of the residency variable, these

minimum numbers of years do not need to be consecutive. Both the SMR and SIR will be adjusted for differences in the age and sex distribution of the populations being compared.

6.6 Comparison of Cancer Rates within Port Colborne

As previously discussed, we will compare rates of respiratory cancer within Port Colborne using three exposure indices: (1) duration of residency in Port Colborne, (2) the average exposure based on soil sample levels and (3) the geographical proximity of the residence to the INCO plant. As detailed in section 6.3, the tabulation of person-years will be used to construct these exposure indices. In order to examine different possible latency effects, in addition to the modelling of a time-dependent exposure index for geographical distance, we will also examine other distance measures:

- Distance to plant from residence lived in longest
- Distance to plant from residence lived in earliest during follow-up
- Distance to plant from residence lived in most recently.

As residential location is based on six-digit postal code, geographical distances will be calculated from the center of the area represented by the postal code to the plant.

As before, Poisson regression will be used to perform risk assessment for these two exposure indices. For the exposure index based on the metal contour levels in soil, the rate ratio will be generated by comparing those in the upper 75% percentile to those in the lowest 25% percentile. For exposure based on geographical distances, Port Colborne cohort members will be grouped into quartiles according to estimated values of these distance measures. Rate ratios will be calculated within each quartile to evaluate whether or not rates of respiratory cancer decrease as distance from the INCO plant increases. Maps will be generated that illustrate respiratory cancer rates across these four quartile regions.

The impact of occupational exposure on presented risk estimates will be evaluated by creating a variable based on dichotomous occupation and residence exposures (see Table **3**). Rate ratios will be estimated by fitting Poisson regression models with this categorical four-level variable. The referent group will consist of those with both low residential and low occupational exposure.

6.7 Adjustment for Selected Confounding Variables

For both referent populations (i.e. comparator communities and Ontario), the regression model will be extended to adjust the rate ratios for the potential confounding influence of family income. This variable is available for each cohort member from the T1FF file. We will evaluate the effect of the number of known years of residency in Port Colborne based on the residential history as documented in the T1FF.

We will adjust the rate ratios for differential rates of smoking that exist between the various populations. This will be done by using survey data collected at a public health unit level. Because these data are collected at a different level (health region versus

community), hierarchical regression or GEE methods will be applied when adjusting for the possible confounding influence of smoking. It is important to recognize that because the outcomes and person-years of follow-up have been divided according to specified age-grouping, sex and calendar period, our rate ratios can also readily be adjusted for the influence of these three factors.

Given that exposure cannot be ascertained prior to 1982 from the T1FF file, age has a strong potential to interact with exposure. Specifically, those cohort members who are older at baseline (1982) have a greater potential for having more years of unmeasured exposure. For this reason, we will also examine an interaction term that consists of exposure and age at entry into the cohort. Further, section 4.1.2 outlines a method for assessing the probability of a Port Colborne cohort member having residence in Port Colborne prior to 1982 in order to determine the extent to which these individuals have the potential for increased exposure.

6.8 Sensitivity Analysis

Both microsimulation models and other adjustments for measurement error will be developed to assess the impact that differential rates of smoking may have on the risk estimates. The simulation models will incorporate various assumptions about the smoking characteristics of the Port Colborne and comparator cohort members. Specifically, plausible differences in the smoking profiles of these two populations will be simulated. These various profiles, in combination with published information about the association between smoking and disease, will be used to evaluate the impact that smoking may have on our presented risk estimates. The nature of this bias will then be incorporated into the risk models to determine a range for possible risk estimates.

7. DATA STORAGE AND TRANSFER

Statistics Canada personnel who have access to the necessary databases will perform the linkage within the T1FF to create the study cohort, and the linkage of the study cohort to the cancer incidence and mortality databases. All personal identifiers (surname and given name) will be stripped from the final electronic database, and summary tables will be calculated for the various research objectives. Some birth date information (birth year and month) will be retained in order to estimate person-years by age grouping. The day of birth will be dropped so that it is not possible to identify a specific Port Colborne resident. The cohort data will be analysed on site at Statistics Canada. Ventana personnel will interact with Statistics Canada employees to analyse the data. The data will be stored in electronic and paper form for five years. Both paper and electronic forms of the data will be maintained in a secure document storage facility. Any transfer of data will be performed as set forth by the guidelines of the Statistics Act. Specifically, assurances must be given to Statistics Canada that the study is valid, the data will be securely stored, and that appropriate steps will be followed such that no individual can be identified using these data.

8. DURATION

It is anticipated that the entire process of protocol approval, data retrieval and analyses, and completion of the final study report will take place through to September 2004. This is the most conservative estimate of study duration since the largest range of Statistics Canada timelines were imputed for these purposes. A number of items outside of Ventana's control will limit the ability of Ventana's resources to expedite the final study report. These items include (1) the submission for database linkage to Statistics Canada can only be initiated following IRB approval of a final protocol design, (2) retrieval of data from Statistics Canada can only follow numerous governmental approvals for their linkage processes, and (3) the time necessary for database linkage by Statistics Canada.

9. PUBLICATION POLICY

The study investigator(s) has(ve) the right to publish, present or otherwise disclose his/her/their findings in the scientific literature with respect to data generated by the Investigator(s) from the study, subject to the following criteria:

- All final draft manuscripts, based on whole or in part on the study, must undergo a review and be approved by the TSC. The sponsor will also be provided with a copy of the publication for response. Submission to the TSC for review must occur at least 30 days prior to submission for publication by the Investigator(s) of any manuscript.
- No more than six months must elapse following completion of the final study report before submission of a first publication.

Should the TSC determine that the manuscript contains incorrect information, or that it discloses confidential or proprietary information, the Investigator(s) will either remove it or modify the manuscript to the satisfaction of the TSC, so that publication of the revised manuscript may proceed. Statistics Canada will have an opportunity to comment on the final draft report and manuscript associated with this study.

All relevant government institutions providing data and assistance with statistical linkage will be appropriately acknowledged in all scientific publications resulting from this work. In addition to scholarly publication, the aggregate results and findings of this study will be made freely available to all interested parties and associated health authorities.

10. REFERENCES

ATSDR, 2001. *Draft Toxicological Profile for Cobalt*. U.S. Department of Health and Human Services. Public Health Service. Agency for Toxic Substances and Disease Registry. Atlanta, GA.

ATSDR, 2000. *Toxicological Profile for Arsenic*. U.S. Department of Health and Human Services. Public Health Service. Agency for Toxic Substances and Disease Registry. Atlanta, GA.

ATSDR, 1997. *Toxicological Profile for Nickel*. U.S. Department of Health and Human Services. Public Health Service. Agency for Toxic Substances and Disease Registry. Atlanta, GA.

ATSDR, 1990. *Toxicological Profile for Copper*. U.S. Department of Health and Human Services. Public Health Service. Agency for Toxic Substances and Disease Registry. Atlanta, GA.

Band PR, Gaudette LA, Hill GB et al. 1993. *The making of the Canadian Cancer Registry: cancer incidence in Canada and its regions, 1969 to 1988.* Health and Welfare Canada, Canadian Council of Cancer Registries, Statistics Canada. Ottawa: Minister of Supply and Services Canada. Cat. No. C52-42/1992.

Band PR, Le ND, Fang R, et al. 2001. Cohort cancer incidence among pulp and paper mill workers in British Columbia. *Scand J Work Environ Health* 27(2):113-9.

Bleuer S. 1996. Gross flows estimation from longitudinal administrative data with missing waves. *Proceedings of the Survey Research Methods Section, American Statistical Association: Sample Attrition in Longitudinal Surveys* 22:681-686.

Blot WJ, Fraumeni JF Jr. 1975. Arsenical air pollution and lung cancer. *Lancet* 2(7926):142-4.

Breslow NE and Day NE (1987). Statistical methods in cancer research. Volume 2 - The design and analysis of cohort studies. IARC Scientific Publication No. 82. International Agency for Research on Cancer, Lyon.

Brown LM, Pottern LM, Blot WJ. 1984. Lung cancer in relation to environmental pollutants emitted from industrial sources. *Environ Res* 34(2):250-61.

Bulbulyan MA, Jourenkova NJ, Boffetta P, et al. 1996. Mortality in a cohort of Russian fertilizer workers. *Scand J Work Environ Health* 22(1):27-33.

Canadian Council of Ministers of the Environment. 1999. Canadian soil quality guidelines for the protection of environmental and human health: Nickel 1999. In: Canadian environmental quality guidelines, 1999, Canadian Council of Ministers of the Environment, Winnipeg.

Cantor KP. 1997. Drinking water and cancer. *Cancer Causes Control* 8:292-308.

Collishaw NE, Tostowaryk W, Wigle DT. 1988. Mortality attributable to tobacco use in Canada. *Can J Public Health* 79:166-9.

Cordier S, Theriault G, Iturra H. 1983. Mortality patterns in a population living near a copper smelter. *Environ Res* 31(2):311-22.

CSTEE, 2001. Opinion on: Position paper on ambient air pollution by nickel compounds. Final version October 2000. Opinion expressed at the 22nd CSTEE plenary meeting, Brussels, 6/7 March 2001.

Doll R, Andersen A, Cooper WC, et al. 1990. Report of the International Committee on Nickel Carcinogenesis in Man. *Scand J Work Environ Health*. 16(1 spec no.):1-82.

Ellison LF, Morrison HI, de Groh M, Villeneuve PJ. 2000. Health Consequences of Smoking Among Canadian Smokers: An Update. *Chronic Diseases in Canada* 20(3):36-9.

Enterline PE, Day R, Marsh GM. 1995. Cancers related to exposure to arsenic at a copper smelter. *Occup Environ Med* 52(1):28-32.

Finkelstein MM. 1996. Clinical measures, smoking, radon exposure, and risk of lung cancer in uranium miners. *Occup Environ Med* 53(10):697-702.

Frost F, Harter L, Milham S et al. 1987. Lung cancer among women residing close to an arsenic emitting copper smelter. *Arch Environ Health* 42(2):148-52.

Gaudette LA, Lee J. 1997. *Cancer incidence in Canada, 1969-1993.* Statistics Canada cat. no. 82-566-XPB edition. Ottawa: Minister of Industry.

Greenland S. 1996. Basic methods for sensitivity analysis of biases. *Int J Epidemiol* 25(6):1107-16.

Hayes RB. 1997. The carcinogenicity of metals in humans. *Cancer Causes Control* 8(3):371-85.

Health Services Department, Region of Niagara, Technical Report: Assessment of Potential Risks of Reported Soil Levels of Nickel, Copper and Cobalt in Port Colborne and Vicinity. 1997, Regional Niagara Public Health Department.

Hertzman C, Teschke K, Ostry A, et al. 1997. Mortality and cancer incidence among sawmill workers exposed to chlorophenate wood preservatives. *Am J Public Health* 87(1):71-9.

Hertz-Picciotto I. 1998. Environmental Epidemiology. Chapter 28 in: Modern Epidemiology, 2nd Edition. (Rothman KJ and Greenland S, eds). Lippincott-Raven Publishers. Philadelphia, USA.

Hogstedt C, Alexandersson R. Mortality study among workers from the hard metal industry (in Swedish). Arbete och Halsa 1990; 21:1-10 as cited in Lison D, De Boeck M, Verougstraete V, Kirsch-Volders M. 2001. Update on the genotoxicity and carcinogenicity of cobalt compounds. 58:619-25.

Howe GR, Lindsay J. 1981. A generalized iterative record linkage computer system for use in medical follow-up studies. *Comput Biomed Res* 14:327-40.

IARC 1980. Arsenic and Arsenic Compounds. *IARC Monographs Eval Carcinog Risks Humans*. Vol. 23. IARC: Lyon, France.

IARC 1987. Overall Evaluations of Carcinogenicity: an updating of IARC monographs 1 to 42. *IARC Monographs Eval Carcinog Risks Humans*. Supp. 7. IARC: Lyon, France.

IARC 1990. Chromium, Nickel and Welding. *IARC Monographs Eval Carcinog Risks Humans*. Vol. 49. IARC: Lyon, France.

IARC, 1991. Chlorinated drinking water; chlorination by-products; some other halogenated compounds; cobalt and cobalt compounds. *IARC Monogr Eval Carcinog Risks Hum.* 52:1-544. International Agency for Research on Cancer Working Group, Lyon, 12-19 June 1990.

Jain MG, Ferrence RG, Rehm JTet al. 2000a. Alcohol and breast cancer mortality in a cohort study. *Breast Cancer Res Treat* 64(2):201-9.

Jain MG, Rohan TE, Howe GR, Miller AB. 2000b. A cohort study of nutritional factors and endometrial cancer. *Eur J Epidemiol* 16(10):899-905.

Järup L, Pershagen G, Wall S. 1989. Cumulative arsenic exposure and lung cancer in smelter workers: a dose-response study. *Am J Ind Med* 15(1):31-41.

Julian JA, Muir DCF. 1996. A study of cancer incidence in Ontario nickel workers. Final report. Toronto: Occupational Disease Panel, January 1996. *as cited in*: Report to the Workers' Compensation Board on Cancer of the Larynx in Workers in Primary Nickel Production, Occupational Disease Panel Report No. 19, July, 1997.

JWEL, 2001. Draft Summary Report on Chemicals of Concern Evaluation; Port Colborne Community Based Risk Assessment.

Katsnelson BA, Kosheleva AA, Kuzmin SV, et al. 2002. Environmental and individual risk factors for lung cancer in dwellers of a small town with unusually high cancer morbidity rates. *Abstract*: 12th

conference of ISEA and 14th Conference of ISEE "Linking exposures and health: innovations and interactions", Vancouver, Canada, August, 2002.

LaBallois T, Wysocki M, Grabowiecki F. 1999. "A Comparison of Direct Match and Probabilistic linkage in the Death Clearance of the Canadian Cancer Registry" in Record Linkage Techniques - 1997: Proceedings of an International Workshop and Exposition, Chapter 7. Committee on Applied and Theoretical Statistics, National Research Council. Federal Committee on Statistical Methodology, Office of Management and Budget. National Academy Press.

Lasfargues G, Wild P, Moulin JJ, et al. 1994. Lung cancer mortality in a French cohort of hard-metal workers. *Am J Ind Med* 26(5):585-95.

Leclerc A, Goldberg M, Goldberg P et al. 1987. Geographical distribution of respiratory cancer in New Caledonia. *Arch Environ Health* 42(5):315-20.

Lison D, De Boeck M, Verougstraete V, Kirsch-Volders M. 2001. Update on the genotoxicity and carcinogenicity of cobalt compounds. 58:619-25.

Lubin JH, Pottern LM, Stone BJ, Fraumeni JF Jr. 2000. Respiratory cancer in a cohort of copper smelter workers: results from more than 50 years of follow-up. *Am J Epidemiol* 151(6):554-65.

Matanoski GM, Landau E, Tonascia J, et al. 1981. Cancer mortality in an industrial area of Baltimore. *Environ Res* 25(1):8-28.

MOE 1997. Guideline for use at contaminated sites in Ontario. Revised February 1997. Queen's Printer for Ontario. Report No. ISBN 0-7778-6114-3.

MOE 2000. *Phytotoxicology Soil Investigation: INCO – Port Colborne (1999).* Queen's Printer for Ontario. Report No. SDB-031-3511-2000.

MOE 2002. *Soil Investigation and Human Health Risk Assessment for the Rodney Street Community: Port Colborne*. Queen's Printer for Ontario. Report No.: SDB-010-3511-2001.

Monson, Richard R. Occupational epidemiology. 2nd ed. CRC Press, 1980. WA950 M7550 1990

Moulin JJ, Wild P, Mur JM, et al. 1993. A mortality study of cobalt production workers: An extension of the follow-up. *Am J Ind Med* 23:281-8.

Moulin JJ, Wild P, Romazini S, et al. 1998. Lung cancer risk in hard-metal workers. *Am J Epidemiol* 148(3):241-8.

Mur JM, Moulin JJ, Charruyer-Seinerra MP, Lafitte J. 1987. A cohort mortality study among cobalt and sodium workers in an electrochemical plant. *Am J Ind Med* 11:75-81.

National Cancer Institute of Canada. 2003. *Canadian Cancer Statistics 2003*. Toronto, Canada.

Norseth T, Piscator M. Nickel. In: Friberg L, Nordberg GF, Vouk VB, editors. Handbook on the toxicology of metals. Amsterdam: Elsevier/North-Holland Biomedical Press, 1979. *As cited in:* Denkhaus E, Salnikow K. 2002. Nickel essentiality, toxicity, and carcinogenicity. Critical Reviews in Oncology/Hematology 42:35-56.

Pershagen G. 1985. Lung cancer mortality among men living near an arsenic-emitting smelter. *Am J Epidemiol* 122(4):684-94.

Prescott E, Netterstrom B, Faber J, et al. 1992. Effect of occupational exposure to cobalt blue dyes on the thyroid volume and function of female plate painters. *Scan J Work Environ Health* 18(2): 101-4.

Qiao YL, Taylor PR, Yao SX, et al. 1997. Risk factors and early detection of lung cancer in a cohort of Chinese tin miners. Ann Epidemiol 7:533-41.

Rom WN, Varley G, Lyon JL, Shopkow S. 1982. Lung cancer mortality among residents living near the El Paso smelter. *Br J Ind Med* 39(3):269-72.

Rothman, Kenneth J. and Sander Greenland. Modern Epidemiology <http://lww.com/cgi-bin/wwonline.storefront/1799503477/Product/View/0&2D316&2D75780&2D2>. 2nd ed. Hagerstown MD, Lippincott-Raven <http://www.lrpub.com/>, 1998. Copies ordered.

Sabbioni E, Mosconi G, Minoia C, Seghizzi P. 1994. The European Congress on Cobalt and Hard Metal Disease. Conclusions, highlights and need of future studies. *Sci Total Environ* 150(1-3):263-70.

SAS Institute, Statistical Analysis System (SAS). 2002: Raleigh, NC.

Statistics Canada. Canadian Cancer Registry. 2003<http://stcwww.statcan.ca/english/sdds/3207.htm>

Statistics Canada. *Health Region Peer Groups*. Ottawa, Statistics Canada, 2002.

Statistics Canada. 2000. *Family Data: User's Guide* Small Area and Administrative Data Division, Statistics Canada. Statistics Canada Product #13C0016.

Statistics Canada. *GRLS V2: generalized record linkage system concepts*. Research and General Systems, Systems Development Division. Ottawa, 1993.

Steenland K, Beaumont J, Halperin W. 1984. Methods of control for smoking in occupational cohort mortality studies. *Scand J Work Environ Health* 10: 143-149.

Swennen B, Buchet J-P, Stánescu D, et al. 1993. Epidemiological survey of workers exposed to cobalt oxides, cobalt salts and cobalt metal. *Br J Ind Med* 50: 835-42.

Terry P, Jain M, Miller AB, et al. 2002. Dietary intake of folic acid and colorectal cancer risk in a cohort of women. *Int J Cancer*. 97(6):864-7.

Thomas DC. 1987. Use of computer simulation to explore analytical issues in nested case-control studies of cancer involving extended exposures: methods and preliminary findings. *J Chron Dis* 40 (Suppl 2):201S-208S.

Tollestrup K, Daling JR, Allard J. 1995. Mortality in a cohort of orchard workers exposed to lead arsenate pesticide spray. *Arch Environ Health* 50(3):221-9.

Tuchsen F, Jensen MV, Villadsen E, Lynge E. 1996. Incidence of lung cancer among cobalt-exposed women. *Scand J Work Environ Health* 22(6):444-50.

U.S. Department of Health and Human Services. 1990. *Smoking and Health: A National Status Report, 2nd Edition. A Report to Congress.* U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Maryland, USA. DHHS Publication No. (CDC) 87-8396.

U.S. Department of Health and Human Services. 1989. Reducing the Health Consequences of Smoking: 25 Years of Progress. A Report of the Surgeon General. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. DHHS Publication No. (CDC) 89-8411.

U.S. Department of Health and Human Services. 1983. *The Health Consequences of Smoking: Cadiovascular Disease. A Report of the Surgeon General.* U.S. Department of Health and Human

Services, Public Health Service, Office on Smoking and Health. Maryland, USA. DHHS Publication No. (PHS) 84-50204.

U.S. Department of Health and Human Services. 1982. *The Health Consequences of Smoking: Cancer. A Report of the Surgeon General.* U.S. Department of Health and Human Services, Public Health Service, Office on Smoking and Health. Maryland, USA. DHHS Publication No. (PHS) 82-50179.

U.S. EPA. 1986. Health Assessment Document for Nickel and Nickel Compounds. Prepared by the Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office, Research Triangle Park, NC. EPA/600/8- 83/012FF.

Ventana Clinical Research Corporation. 2003a. *An Overview of Available Data Sources and Possible Epidemiologic Study Designs to Evaluate Patterns of Cancer Incidence and Mortality in Port Colborne.* Toronto, Canada.

Ventana Clinical Research Corporation. 2003b. *Protocol C Technical Methodology Document: The Selection of Ontario Comparator Communities for Port Colborne.* Toronto, Canada.

Ventana Clinical Research Corporation. 2002. *Overview of Proposed CHAP Research in Port Colborne.* Toronto, Canada.

Ventana Clinical Research Corporation. 2001. *Draft Scope of Work: Port Colborne Community Health Assessment Project.* Toronto, Canada.

Wall S. 1980. Survival and mortality pattern among Swedish smelter workers. *Int J Epidemiol* 9(1):73-87.

Welch K, Higgins I, Oh M, and Burchfiel C. 1982. Arsenic exposure, smoking, and respiratory cancer in copper smelter workers. *Arch Environ Health* 37(6):325-35.

Williams FLR, Ogston SA. 2002. Identifying populations at risk from environmental contamination from point sources. *Occup Environ Med* 59(1):2-8.

Wingren G, Axelson O. 1993. Epidemiologic studies of occupational cancer as related to complex mixtures of trace elements in the art glass industry. *Scand J Work Environ Health* 19(suppl. 1):95-100.

WHO Regional Office for Europe, Copenhagen, Denmark. 2000. Air Quality Guidelines – Second Edition: Chapter 6.10 Nickel.

Wong O, Whorton MD, Foliart DE, Lowengart R. 1992. An ecologic study of skin cancer and environmental arsenic exposure. *Int Arch Occup Environ Health* 64(4):235-41.

Xu Z-Y, Blot WJ, Xiao H-P, et al. 1989. Smoking, air pollution, and the high rates of lung cancer in Shenyang, China. *J Natl Cancer Inst* 81(23):1800-6.

APPENDIX I Maps

APPENDIX II Comparator Communities

The following table lists the communities with small Mahalanobis distances to Port Colborne based on discriminant analysis of 47 predictor variables.

No.[a]	Census Subdivision Number	Census Subdivision Name	Mahalanobis Distance to Port Colborne	T1FF Coverage Rates (%)[b]
-	3526011	Port Colborne	0.00000	101.8
G	3526047	Niagara-on-the-Lake	1.74079	78.0
C	3559046	McCrosson and Tovell	2.22782	—
C	3547039	Eganville	2.67320	335.8
G	3526043	Niagara Falls	2.94220	100.3
1	3554056	Matachewan	2.98439	123.4
C	3537022	Harrow	3.01319	328.1
G	3526032	Welland	3.25407	107.1
2[E]	3557061	Sault Ste. Marie	3.38805	108.2
3	3537006	Leamington	3.51237	95.2
G	3526037	Thorold	3.62311	91.1
G	3553007	Sudbury	4.00671	81.9
C	3559041	Atwood	4.16355	719.4
C	3547036	South Algona	4.48652	—
4	3554068	Kirkland Lake	4.60732	94.6
G	3526003	Fort Erie	4.67671	56.5
C	3512048	Tudor and Cashel	4.80399	—
C	3534036	West Lorne	4.81849	203.7
C	3547028	Barry's Bay	5.12654	275.6
C	3547096	Deep River	5.22742	130.6
5[E]	3558004	Thunder Bay	5.39013	108.3
C	3537039	Windsor	5.48839	125.5
C	3539004	Wardsville	6.03647	199.0
C	3549036	Carling	6.08293	—
C	3547026	Sherwood Jones and Burns	6.25068	—
C	3547012	Griffith and Matawatchan	6.25877	—
C	3539016	Strathroy	6.43790	125.9
6	3551001	Tehkummah	6.45543	119.9
C	3560041	Red Lake	6.52501	55.3
G	3526057	Lincoln	6.55337	—
C	3554058	McGarry	6.64969	—
G	3526053	St. Catharines	6.65532	101.9
C	3549056	South River	7.08151	213.5
C	3537004	Mersea	7.09821	—
C	3537009	Gosfield South	7.15478	—
C	3549005	The Archipelago	7.34054	—

No.[a]	Census Subdivision Number	Census Subdivision Name	Mahalanobis Distance to Port Colborne	T1FF Coverage Rates (%)[b]
G	3552031	Nairn	7.36860	—
C	3515011	North Monaghan	7.40302	—
C	3525003	Stoney Creek	7.55926	99.8
G	3526065	Grimsby	7.71629	96.5
C	3547031	Hagarty and Richards	7.96175	—
C	3529009	Oakland	8.12783	23.9
C	3548001	Airy	8.13318	—
C	3528049	Delhi	8.19646	45.0

[a] G=Exclusion due to geographical proximity to Port Colborne; Sudbury exclusion due to similarity in community environmental and occupational exposures; Nairn exclusion due to geographical proximity to Sudbury.

[a] E=A variable will be coded ('#^{E}'=1, '#'=0) for communities with possible exposures to environmental contaminations as indicated by the Ministry of Environment for inclusion in regression analyses.

[a] C=This community will be excluded from the listing of comparator community for this study due to conversion constraints of using postal code data to represent census communities.

[b] Coverage rates are a ratio of the T1FF 2000 population estimates to the 2001 Canadian census estimates

Source: Ventana, 2003

APPENDIX III Hierarchy of Exposure Assessment

Level	Exposure	Variable Type	Level of Detail	Analysis	Comparison
1	Port Colborne Residence	Categorical	Number of years resident in Port Colborne or comparator communities as defined in the T1FF.	*Primary analysis for* 1. primary objective; to compare incidence and mortality rates of respiratory cancer 2. secondary objective; to compare rates for all cancers and causes of mortality	*External* Port Colborne resident (exposed) vs. comparator community resident (unexposed); Ontario
		Occupationally exposed excluded	Resident in Port Colborne and NOT an employee of INCO or resident in comparator communities	*Sub-analysis for* 1. assessing rates observed with primary objective; to compare incidence and mortality rates of respiratory cancer 2. assessing rates observed with secondary objective; to compare rates of all cancer and causes of death	*External* Port Colborne resident (exposed) vs. comparator community resident (unexposed); Ontario
2	Average Annual CoC Exposure	Categorical	Number of years at a particular location (i.e., postal code) combined with an estimate of the soil CoC levels at that location, accumulated over a residential history in Port Colborne, as defined in the T1FF (Figure 1) or similarly defined residence in a comparator community.	*Secondary analysis for* 1. assessing rates observed with primary objective; to compare incidence and mortality rates of respiratory cancer 2. assessing rates observed with secondary objective; to compare rates of all cancers and causes of death 3. exploring associations with CoC exposure by incorporating regional variations in potential exposure; to compare rates of respiratory cancer, all cancers and causes of death	*External* [a] Port Colborne resident within 75th percentile E_{avg} (exposed) vs. Ontario resident (unexposed) Port Colborne resident within 25th percentile E_{avg} (exposed) vs. Ontario resident (unexposed) *Internal Cohort* [b] Port Colborne resident within 75th percentile E_{avg} (exposed) vs. comparator community resident (unexposed) Port Colborne resident within 25th percentile E_{avg} (exposed) vs. comparator community resident (unexposed) Port Colborne resident within 75th percentile E_{avg} (exposed) vs. Port Colborne resident within 25th percentile E_{avg} (unexposed)

Level	Exposure	Variable Type	Level of Detail	Analysis	Comparison
		Occupational exposure incorporated	Number of years at a particular location (i.e., postal code) combined with an estimate of the soil CoC levels at that location, accumulated over a residential history in Port Colborne AND occupational status at INCO, as defined in the T1FF or similarly defined residence in a comparator community (Table 3).	*Secondary analysis for* 1. assessing rates observed with primary objective; to compare incidence and mortality rates of respiratory cancer 2. assessing rates observed with secondary objective; to compare rates of all cancers and causes of death 3. exploring associations with CoC exposure by incorporating regional variations in potential exposure; to compare rates of respiratory cancer, all cancers and causes of death	*Internal Cohort* First category of Port Colborne resident/INCO employee as referent Each of four categories of Port Colborne resident/INCO employee (exposed) (Table 3) vs. comparator community resident (unexposed)
3	Geographical Proximity to INCO Plant	Categorical	Categories (grouped by quartile) based on residence lived longest, lived first and lived last	*Secondary analysis for exploring:* Associations between CoCs and incidence rates within Port Colborne	*Internal Cohort* Rates within each quartile defined by geographical proximity

[a] Sample size permitting. Additional analyses will be performed with alternative categorizations of E_{avg} (e.g., 'low' < 75th percentile; 'high' ≥ 75th percentile) to further define those with potential increased exposures.

[b] Sample size permitting.

APPENDIX IV Causes of Death (ICD-9 Codes)

Causes of Death	ICD-9 Code
Infectious and parasitic diseases	001 – 139
Cancer	140 – 208
Disorders of thyroid gland	240 – 246
Psychoses	290 – 299
Central nervous system	320 – 349
Peripheral nervous system	350 – 359
Heart disease	393 – 429
Cerebrovascular disease	430 – 438
Diseases of arteries, arterioles and capillaries	440 – 448
Respiratory system	460 – 519
Digestive system	520 – 579
Genitourinary system	580 – 629
Skin and subcutaneous tissue	680 – 709
Musculoskeletal system and connective tissue	710 – 739
Congenital anomalies	740 – 759
Injury and poisoning	800 – 999
Motor vehicle traffic accidents	E810 – E819
Accidental falls	E880 – E888