MISCLASSIFICATION OF EXPOSURE: COFFEE AS A SURROGATE FOR CAFFEINE AND METHYLXANTHINE INTAKE

by

Janet Brown

A thesis submitted in conformity with the requirements
for the degree of Master of Science
Graduate Department of Community Health
University of Toronto

MISCLASSIFICATION OF EXPOSURE: COFFEE AS A SURROGATE FOR CAFFEINE AND METHYLXANTHINE INTAKE. Janet Brown, Master of Science, Graduate Department of Community Health, University of Toronto, 1997.

This cross-sectional study obtained, by mail, 481 self-administered questionnaires from a sample of men and women aged 30-75 years from Southern Ontario. Questions were designed to assess daily and lifetime intake of coffee, caffeine and methylxanthine (MTX) in foods, beverages and medications. Exposure estimates were used to determine misclassification which occurs when coffee is used as a surrogate measure for caffeine and MTX consumption. The effect of misclassification was studied in two ways: preselecting odds ratios (ORs) by constructing a hypothetical case-control distribution, and "correcting" OR estimates reported in the literature. The kappa statistic revealed coffee as a poor measure of caffeine and MTX intake for daily and lifetime intake. OR corrections for multiple levels of exposure showed that using coffee as a surrogate measure for caffeine would mask any true association regardless of age, sex, and dose-response for daily and lifetime exposures.

I would like to thank my supervisor for introducing me to this interesting topic and for allowing me the freedom to explore it on my own and my thesis committee members, Gerarda Darlington and Jack Williams, for their ongoing support. I would also like to thank Peggy Sloan for always being available to answer questions, Don Maclean and Joe Downey for helping me with the seemingly endless task of subject recruitment, Gayle Morrison and Lori-Ann Larmand for doing the data entry and Darlene Zimmerman who made editing effortless.

LIST OF FIGURES

1. INTRODUCTION

Despite extensive examination in epidemiologic studies, coffee and caffeine consumption have not been consistently associated with disease occurrence. In some areas of research, coffee has been the focus of interest (e.g., pancreatic cancer, cardiovascular disease) while in other areas, caffeine has been more intently studied (e.g., breast disease, reproductive issues). 'Coffee' and 'caffeine' are used almost interchangeably in the literature, despite the lack of evidence that they are equivalent exposures. This lack of distinction has led to incomplete caffeine measurement: using only coffee or tea and failing to include other significant sources of caffeine (Stavric *et al.*, 1988; Bullough *et al.*, 1990; Pozniak, 1985). An additional concern is the method of measurement; researchers often fail to account for cup volume or brewing method, both of which affect caffeine content.

Caffeine is one of a group of closely related chemicals called methylxanthines (MTXs), theobromine and theophylline complete the group. Theobromine is found in tea, chocolate and cocoa; theophylline is found in tea and prescribed for use as a bronchodilator for patients with asthma and chronic bronchitis. Significant amounts of caffeine are contained in regular tea, cola soft drinks and some medications, as well as in coffee.

The effects of MTXs (caffeine, theobromine and theophylline) have been studied together with respect to some diseases, such as benign breast disease (e.g., Rohan *et al.*, 1989; Bullough *et al.*, 1990) and breast cancer (e.g., Lubin & Ron, 1990; McLaughlin *et al.*, 1992). However, these studies have not included MTX contained in medications. With respect to pancreatic cancer, coffee intake has been studied, with few attempts at measuring caffeine intake and no attempt at studying MTX (Gordis, 1990). Because caffeine, theobromine and theophylline have similar structures, metabolism, and effects, it seems reasonable to study these three sources of MTX together. If coffee, caffeine and MTX intake do not represent equivalent exposures, then studies that omit sources of MTXs other than coffee may preclude the identification of a potential association between MTX and disease status.

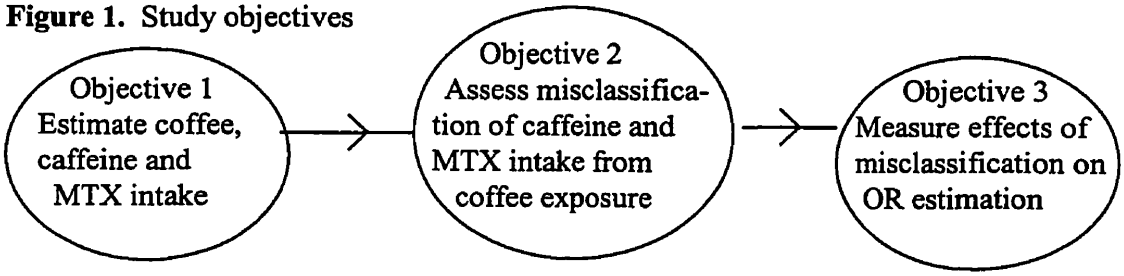
There is extensive literature on the effects of measurement error (e.g., Biemer *et al.*, 1991; Armstrong *et al.*, 1992) and misclassification of exposures (e.g., Marshall, 1994; Lyon, 1992; Birkett, 1992), which is relevant to the estimation of the relative risk associated with MTX exposure. The specific error, that of not including all sources of MTX, is assumed to be independent of disease status and, thus, non-differential. As a consequence, an unadjusted OR estimates would be expected to be biased toward the null hypothesis of no effect for dichotomous exposure (Kleinbaum *et al.*, 1982; Biemer *et al.*, 1991). The outcome is less clear for multiple levels of exposure (Birkett, 1992).

Because exposure to MTX is so prevalent in the general population, even a small excess risk of disease related to MTX intake would constitute a substantial attributable risk.

1.1 Objectives

This study assesses daily and lifetime intake of coffee, caffeine and MTX from a sample of the population in Southern Ontario (see Figure 1). The exposure estimates were used to assess the misclassification, always an underestimation of true exposure, which may occur when coffee is used as a surrogate measure for caffeine and for total MTX consumption. These misclassification estimates provided the basis for studying the effects of underestimation of risk by two methods. The first method involved preselecting odds ratios (ORs) by constructing a hypothetical case-control distribution using true measures of caffeine intake and comparing the resulting OR estimates when coffee is used to approximate caffeine intake. This approach was general and can be applied to all diseases. The second method involved "correcting" OR estimates related to pancreatic cancer reported in the literature. This area was chosen because the findings are particularly inconsistent.

Figure 1. Study objectives

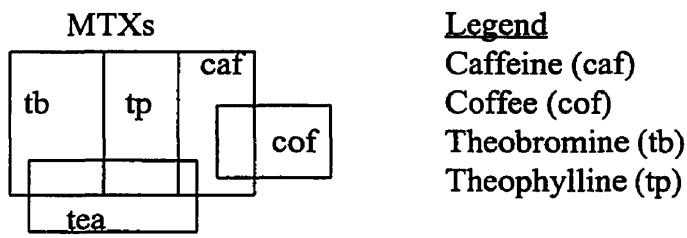


2.1 Overview of the Methylxanthines (MTXs)

2.1.1 Description and effects

Caffeine and other methylxanthines (MTXs) form a group of alkaloid chemicals which are structurally similar to two nucleic acids, adenine and guanine. In addition to caffeine, the health effects of two other MTXs that have been studied are theobromine, found in tea and chocolate, and theophylline, found in tea and asthma medications (see Figure 2).

Figure 2. The inter-relationship between the MTXs, coffee and tea.



Theophylline (1,3-dimethylxanthine) and theobromine (3,7-dimethylxanthine) differ from caffeine (1,3,7-trimethylxanthine) only in the number and placement of the methyl groups and possess many of the same drug actions (Ensminger *et al.*, 1994;Yesair *et al.*, 1985). Dietary MTXs are metabolized through similar reactions and produce the same major by-products (Yesair *et al.*, 1985). In addition, theophylline and theobromine are two of the

main metabolites of caffeine, although there is variability in the pathways by which caffeine is metabolized (James, 1991).

All three MTXs are physiologically active. Some of the characteristics of MTXs are their ability to inhibit phosphodiesterases, block adenosine receptors, and mobilize calcium (Debry, 1993; Yesair *et al.*, 1985). Other important effects of MTXs are their inhibition of prostaglandins, histamine and leukotrienes, and their mutagenic and antimutagenic properties. These mechanisms may have important effects in, for example, renal function, the gastro-intestinal and immune systems, osteoporosis, cancer initiation and/or promotion, and in regulation of hormones and blood pressure (Curatolo *et al.*, 1983; James, 1991; Stavric, 1992; Thompson, 1988; Watson, 1988). While similar in many ways, the three MTXs can act differently; caffeine has the most potent effect on the brain and skeletal muscles, theophylline is the most powerful stimulant for the heart, bronchi and kidneys, and theobromine functions as a weak stimulant relative to theophylline and caffeine (James, 1991; Lamarine, 1984; Hirsh, 1985).

Data sources

Caffeine is one of the most widely consumed drugs in North America (Gilbert, 1984; Watson, 1988) and its use has remained fairly consistent in Canada over the past three decades (Garattini, 1994). The only recent Canadian caffeine consumption data available were from two sources: the Coffee Association of Canada (1995 data) and the Ontario Health Survey (1990 data). The Coffee Association of Canada annually surveys 6,000 individuals judged as representative of the Canadian population (Wilks, 1996). Participants keep a diary in which they record coffee consumption for seven consecutive days. The 1990 Ontario Health Survey was an in-depth survey of health status and health behaviours. Information was collected for 63,663 individuals in 28,145 randomly selected households across Ontario (Ministry of Health, 1990). The survey collected information on the coffee and tea consumed in the month prior to the study. No sources of information are available for trends of theobromine or theophylline intake.

According to the Coffee Association, 50% of Canadians drink coffee; 75% of these individuals drink at least one cup per day and the average coffee drinker consumes 3 cups per day (Wilks, 1996). According to the Ontario Health Survey, slightly more than 75% of men and women aged 30-75 drank coffee during the month prior to the survey. For individuals who drank coffee daily, the mean consumption was 1.9 cups. Men and women reported similar intakes but individuals aged 60-75 drank slightly less coffee than other age groups. Approximately half of the men and two thirds of the women drank tea in the month prior to the study. For individuals who drank tea daily, the mean consumption was 1.3 cups and was similar by age and sex, although women aged 60-75 had a slightly higher tea consumption (average of 1.5 cups).

Statistics Canada's data are based on estimates of coffee "disappearance", that is, per capita consumption based on the amount of coffee sold in the country. In 1984, Statistics Canada estimated that the average daily caffeine intake was 238 mg (Gilbert, 1984): 55% was obtained from coffee, 32% from tea, 7% from soft drinks, and 1% from chocolate. The remaining 5% was obtained from medicines and other beverages such as maté -- a drink popular in South America.

International comparisons of intake

Results from a survey done in 1993 in the United States showed that decaffeinated coffee accounted for about 15% of all cups of coffee (Barone & Roberts, 1996). Of all caffeinated coffees, 14% were instant. In the United States, a higher percentage of caffeine intake is from coffee (75%) than in Canada (55%). In the U.S., tea accounted for 15% and soft drinks for 16% of caffeine intake (Hui, 1994). Caffeine intake increases with age until age 60, and at this age tea becomes the major caffeine-containing beverage; it is impossible to determine from these cross-sectional data whether this is a cohort effect or an age-related trend. American women generally consume more caffeine than American men, except for men aged 40 to 49 who consume more caffeine than any other age group of either sex (Pozniak, 1985). Caffeine consumption is much higher in some European countries than in North America. For example, in Britain, average daily per capita use of caffeine is about 445 mg as compared to 238 mg in Canada. In Britain, 72% of caffeine is obtained from tea and 19% from coffee (Gilbert, 1984).

Factors associated with intake

In addition to age and sex, the other best established relationships are the positive associations between coffee drinking, cigarette smoking and alcohol consumption

showed that drinkers of caffeinated coffee drank more alcohol, consumed more dietary saturated fats and cholesterol, were more likely to be current smokers and less likely to be current exercisers than were non-coffee drinkers. Smoking and lack of exercise showed a dose-response relationship to the amount of caffeinated coffee consumed.

2.2 MTX measurement

2.2.1 Variability of MTX content

Coffee and tea

A wide variety of MTX content is apparent among products, coffee having the highest variability of caffeine content among foods and beverages (Barone & Roberts, 1996). The caffeine content of coffee depends, among other things, on the method of preparation. Increasing concentrations of caffeine are found in instant (60-85 mg/cup), brewed (80-135 mg/cup), and percolated methods (115-175 mg/cup) (Chou, 1992). The amount of caffeine in coffee also depends on the species of the coffee bean used and on the darkness of the roast. There are over 20 species of coffee beans; the two main beans

used in North America are the Robusta and the Arabica. Robusta beans have twice the concentration of caffeine of the Arabica (D'Amicis & Viani, 1993; Spiller, 1985).

Tea contains all three MTXs, with caffeine accounting for the largest proportion (30-70 mg/cup). Tea is the only food or beverage source for theophylline (Bullough *et al.*, 1990). Caffeine content varies in tea as well as in coffee. The amount of caffeine and theobromine in tea brewed for 5 minutes is twice the amount of that brewed for 1 minute (Graham, 1984; Stavric *et al.*,1988); levels of theophylline were not reported.

A survey conducted in Ottawa found that although the typical volume of coffee consumed at home and at work was higher (224 and 234 ml) than in commercial establishments (171 ml), commercial establishments tended to brew much stronger coffee (Stavric *et al.*, 1988). The result was approximately equal caffeine intake per typical volume in each locale: the average amount of caffeine in a cup of coffee prepared in the home was 79 mg (range 37-114 mg) and in commercial establishments was 75 mg (range 48-156 mg). What is striking is the degree of variation of caffeine content per cup among subjects, which varied by up to 3 times, regardless of locale.

Cola soft drinks have a caffeine content ranging from 30 mg (e.g., Canada Dry Cola) to 100 mg (e.g., Jolt) per can (Ensminger *et al.*, 1995, National Soft Drink Association, 1993). Coca-Cola contains 65 mg and Pepsi-Cola contains 43 mg of caffeine per can. Non-cola soft drinks, such as root beer or Sprite, do not contain any MTX.

Chocolate products

Chocolate is the main source of theobromine. The range of theobromine content in chocolate is quite large, with higher amounts of theobromine found in types of chocolate which contain higher concentrations of cocoa (e.g., bittersweet as opposed to milk chocolate). The values of theobromine content by brand of hot chocolate range from 30 to 130 mg per cup (De Planer, 1989; Zoumas *et al.*, 1980). Values for chocolate milk range from 35-99 mg per cup (Zoumas *et al.*, 1980). The values of caffeine content in chocolate are usually about one eighth that of theobromine.

Medications

Various drugs supply significant amounts of both caffeine and theophylline. Caffeine is contained in some analgesics (e.g., migraine medications, 222's), menstrual medications (e.g., Midol), and cough, cold and allergy products (e.g., Dristan). Significant amounts of theophylline are contained in asthma and bronchial medications: for example, Uniphyl, Theophylline by injection and Theodur. Each type of medication contains at least 100 mg of theophylline, and some contain more depending on the dosage (CPS, 1995). The contribution of MTX in medications is rarely included in MTX measurement because the use of such medications is considered to be minimal.

It is extremely difficult to develop a single value of MTX for any food or beverage. James (1991; p. 46) writes "No strong case can be made for recommending the use of one or the other of sets of 'standard values' that have been recommended previously".

2.2.2 Coffee and caffeine studies

General measurement error

Lack of precision in the measurement of caffeine intake has been established as a significant source of error in caffeine-related research (Barone & Roberts, 1996; James, 1991; Lamarine, 1984). The measurement of caffeine typically involves estimating the amount of caffeine content in different foods, multiplying it by the usual number of servings, and summing the total over the sources. The reliability of such estimates has been brought into question due to the many factors affecting the amount of caffeine contained in foods. Prior studies have been criticised for ignoring the variations inherent in the product consumed, not distinguishing between the use of caffeinated and decaffeinated coffee, and not differentiating between the effect of caffeine and the effect of other ingredients in coffee (Debry, 1993; James 1991). However, caffeine has been the primary ingredient of interest because it is highly active (see earlier section on effects of MTXs, p.5).

The failure to separate the effects of coffee and caffeine may be due to convenience and could explain why the terms are used interchangeably. Although it has been acknowledged that "the number of cups of coffee or tea consumed daily is frequently

index of intake" (Stavric *et al.*, 1988), the effects of using this index have not been examined.

Coffee/caffeine/MTX debate: what should be studied?

To help assess the potential inadequacy of using coffee as a surrogate for caffeine, it is useful to examine the literature in a particular area, to see the effect of incomplete measurement of caffeine on study results. In the case of pancreatic cancer, it may be that coffee has been somewhat arbitrarily chosen over caffeine. MacMahon *et al.* (1981) found an unexpectedly strong association between coffee and pancreatic cancer, and these results were probably responsible for stimulating interest in coffee. In fact, the mechanisms that might implicate coffee/caffeine in the etiology of pancreatic cancer remain largely obscure (Fredholm, 1984; Stavric *et al.*, 1992). It has been reported that caffeine, caffeinated coffee, and decaffeinated coffee all stimulate various pancreatic secretions, although not in identical ways (Coffey *et al.*, 1986). Coffey *et al.* (1986) also suggested that chronic stimulation of pancreatic secretions may make the pancreas susceptible to carcinogenesis. In summary, what little evidence there is suggesting that coffee may be a risk factor for pancreatic cancer, might apply equally to caffeine.

than caffeine alone, has only been hinted at in the pancreatic cancer literature. For example, Gordis (1990) reviewed the literature, in a paper entitled "Consumption of Methylxanthine-containing Beverages and Risk of Pancreatic Cancer". However, what is interesting and perhaps somewhat revealing about the lack of clarity on this issue is the fact that none of the studies Gordis reviewed actually measured anything more than coffee and tea intake.

2.2.3 Case-control studies on pancreatic cancer

In the area of pancreatic cancer, some studies yielded positive associations with coffee consumption (MacMahon *et al.*, 1981; Mack *et al.*, 1986; Clavel *et al.*, 1989; Lyon *et al.*, 1992) and others no association (Wynder *et al.*, 1983; Olsen *et al.*1989; Falk *et al.*, 1988; Farrow & Davis, 1990; Ghadirian *et al.*, 1991, Jain *et al.*, 1991). Some discrepancies might result from poor diagnosis (Stavric *et al.*, 1988) or from poor choice of controls (Gordis, 1990).

The landmark study by MacMahon *et al.* (1981) found a strong association between coffee consumption and pancreatic cancer. Researchers collected information on the number of cups of tea and coffee consumed in a typical day before illness. For males, each level of coffee exposure (1-2, 3-4, 5+ cups) was associated with an increased risk as compared to the non-exposure level (0 cups), but in a *flat* dose-response relationship (OR estimates = 2.6, 2.3, 2.6 respectively). For females, each level of exposure was associated with an increased risk as compared to the non-exposure level, and in a *positive* dose response relationship (OR estimates = 1.6, 3.3, 3.1 respectively). MacMahon *et al.* (1981) did not distinguish between caffeinated and decaffeinated coffee but argued that decaffeinated coffee had only recently become popular on a large scale, and thus could not be an important contributing factor. Since then, decaffeinated coffee has become increasingly popular (James, 1991). Tea was also measured but was analysed separately from coffee. There was a slight inverse association between pancreatic cancer and tea for men and women. A subsequent study by the same authors did not confirm these results (Hsieh *et al.*, 1986). Nevertheless, the original study instigated a flurry of studies which examined coffee as a variable of interest in pancreatic cancer.

Lyon *et al.*, (1992) conducted a more detailed assessment of lifetime intake of caffeine, measuring intake of coffee (both caffeinated and decaffeinated), tea and soft drinks. Daily coffee intake was associated with an increased risk (OR=1.4) for intake of one to

three cups per day and with an increased risk (OR= 2.3) for four or more cups using non-users as the referent group. Lifetime coffee intake was associated with increased risk which attained significance for males but not females. Coffee was the only variable that was associated with pancreatic cancer. Total lifetime cups of coffee was more strongly related to pancreatic cancer than total caffeine from coffee. Calculations of caffeine from coffee accounted for differing caffeine content by coffee type and mode of preparation. Therefore, the authors concluded that coffee (both caffeinated and decaffeinated) and not caffeine was the exposure of interest. They did not, however, sum caffeine across all sources to include tea and soft drink intake.

Only two studies on coffee and pancreatic cancer have been conducted in Canada, one in Toronto, Ontario (Jain *et al.*, 1989) and the other in Montréal, Québec (Ghadirian *et al.*, 1990). Both studies used the same protocol as part of the SEARCH programme of the International Agency for Research on Cancer. The investigators examined dietary factors, including lifetime intake of coffee and tea, as potential risk factors for pancreatic cancer. Coffee and tea were analysed separately based on quartiles or quintiles of consumption. Coffee was divided into quartiles (in Jain *et al.*, 1989) or quintiles (in Ghadirian *et al.*, 1990) for each coffee type: regular, decaffeinated, ground, ground decaffeinated, instant and instant decaffeinated. No association with pancreatic cancer was found in either study for total coffee, individual coffee subtypes or for tea.

In summary, links between coffee and pancreatic cancer have been frequently hypothesized but not consistently shown. Some studies have measured tea as well as coffee (e.g., MacMahon *et al.*, 1981; Ghadirian *et al.*, 1990; Jain *et al.*, 1989). The only study which examined any other caffeine source was one by Lyon *et al.* (1992) which included cola soft drinks. No other study included caffeine intake from medications and chocolate products.

2.3 Misclassification of exposure

Incomplete exposure assessment for caffeine occurs when coffee intake is used to approximate caffeine intake, given that other potentially important sources of caffeine, such as tea and soft drinks, are omitted. This type of misclassification would likely result in underestimation. For example, if a significant number of non-coffee drinkers acquire caffeine from sources other than coffee, these individuals would be misclassified as non-exposed.

If MTX is the variable of interest then theobromine and theophylline also need to be measured. However, without any data available on theobromine and theophylline intake in the population it is difficult to estimate what percentage of MTX exposure is being overlooked.

The question examined here is "how good an indicator is coffee for caffeine and MTX intake?". Although there has been debate about the uses of the kappa statistic (e.g., Guggenmoos-Holzmann, 1993; Kraemer & Blooch, 1988; Thompson & Walter, 1988), it has been used extensively to estimate the agreement between two measures. Brenner & Kliebsch (1996) and Graham & Jackson (1992), suggest that using weighted kappa corrects for some of the problems arising with multiple categories.

An area that has received little attention is how kappa values are affected by the number of categories. Brenner & Kliebsch (1996) varied the numbers of categories from 2 and 8, and found that the weighted kappa coefficient tends to increase with the number of categories until it levels off at 5 or more categories.

2.3.2 Effect on the odds ratio estimates

Because epidemiology focuses on associations with disease, the impact of measurement error on relative risk estimates or odds ratio estimates is of great importance. Non-differential error in a dichotomous variable tends to bias OR estimates toward one (Willett, 1990; Wacholder, 1995; Armstrong *et al.*, 1994). Under certain circumstances

misclassification can produce an overestimation of relative risk (Copeland *et al.*, 1977), for example, when the misclassification is extreme (over 50%) and the exposed individuals are much more likely to be misclassified as non-exposed than vice-versa (Brenner, 1991; Birkett, 1992), as may be the case for caffeine estimates based only on coffee intake. The effect of misclassification on OR estimation depends on sensitivity and specificity, disease frequency and exposure frequency.

The effect of misclassification on multiple levels of exposure is far more complex. Marshall *et al.* (1990) examined algebraically the relationship between misclassification and the OR estimates, assuming that the misclassification occurred between adjacent exposure categories and that misclassification rates were similar among exposure levels. Their results suggested that even with substantial misclassification, OR estimates are biased toward the null and neither reverse direction nor distort relations. Birkett (1992) extended the conditions of Marshall *et al.* (1990) using different misclassification probabilities among exposure categories and different distributions of subjects among the exposure levels. Based on algebraic models, Birkett showed that the misclassified OR estimates for the highest exposure level will be biased toward the null but that OR estimates for intermediate levels of exposure could be biased in either direction. He found that the amount of bias was influenced by the misclassification rates and by the distribution of subjects among exposure levels.

Dosemeci *et al.* (1990) constructed examples based on data from previous studies and found that misclassification led to several scenarios in multiple levels of exposure. Some of the misclassified OR estimates were biased away from the null, and were biased so as to magnify or reverse the dose-response relationship.

In order to address the effects of using coffee as a surrogate measure for caffeine or MTX intake, a complete assessment of caffeine and MTX intake of a sample population needs to be undertaken. Once true caffeine intake is assessed, it is possible to assess the underestimation that may occur when assuming coffee to be the sole source of caffeine and to measure the effects of this assumption on OR estimates.

3. MATERIAL AND METHODS

3.1 Study sample

Sampling frames for this study came from listings in telephone directories from four regional municipalities located near Metropolitan Toronto. Toronto was not included in the study for a number of reasons including: more relocations among the population result in many out-of-service telephone listings, and the busy city lifestyle may mean fewer answered calls and a lower participation rate (Statistics Canada, 1996). The regions surrounding Toronto, including Durham, Halton-Peel, Hamilton-Wentworth and York, were chosen to provide a wide spectrum of urban, suburban and rural settings of Southern Ontario. This population was also inexpensive to reach and easily contacted for follow-up.

For each of the four regional municipalities, a proportional number of residential listings was randomly selected from the appropriate telephone directory. The four directories totalled 2,457 pages with 1,355,850 listings. The Durham directory contained 244,375 listings, Hamilton-Wentworth 319,574, Halton-Peel 494,343 and York 297,473 listings. Therefore, a sampling fraction of 18.02% was applied to the sample in Durham, 23.57% in Hamilton-Wentworth, 36.46% in Halton-Peel, and 21.94% in York. A random list

without replacement was generated using SAS® to identify the page, column and row of the telephone listing to be called for each telephone directory.

This study included men and women aged 30 - 75 years who were residents of Ontario. The age range was chosen to cover the adult years commonly used in case-control studies measuring coffee/caffeine intake in relation to various diseases. The sample was stratified by sex and age group (i.e., 30-44, 45-59, 60-75) with age groupings chosen to coincide with differences reported previously among exposure levels of coffee and caffeine intake (Pozniak, 1985).

As existing prevalence figures for MTX consumption were not available for a Canadian population, the prevalence reported for a sample of Australian women was used to estimate the sample size (Rohan *et al.*, 1989). A precision of mean MTX intake within ± 5% (α=0.05) by sex and ± 10% (α=0.05) in each of the six age-sex specific strata required a total sample size of 450, 75 in each age-sex stratum. At the time of the telephone call, the potential participants were asked their age in order to keep track of the numbers of participants needed to fill each age-sex quota. Assuming a response rate of approximately 80%, once the individual was eligible and agreed to participate, 564 subjects (94 in each stratum) were required.

3.2 Data collection

Subject recruitment took place from June to November 1995. Up to five initial contact calls were attempted on different days and at different times to maximize the chances of reaching a household member. Calls were placed a maximum of twice on a weeknight (6:30 - 9:30 p.m.), once on a Saturday (10 a.m. - 6 p.m.), once on a Sunday (4 - 9 p.m.) and once during a weekday (10 a.m. - 6:30 p.m.).

The initial contact call established whether the person answering the telephone was eligible and willing to participate (see Appendix A for the scripts). If he or she was not eligible, the interviewer asked to speak to an eligible person. Only one eligible person from each household was asked to participate. During the last month of recruitment, agreement on behalf of another household member was accepted due to time constraints. Consenting subjects were sent a self-administered questionnaire (see Appendix B), together with an explanatory letter (see Appendix C) within two days of the agreement to participate. A stamped, self-addressed envelope was enclosed for the return of the completed questionnaire.

Initially, reminder calls were placed ten days after the questionnaire was sent, but the time frame was lengthened to twelve days when it became apparent that the bulk of the

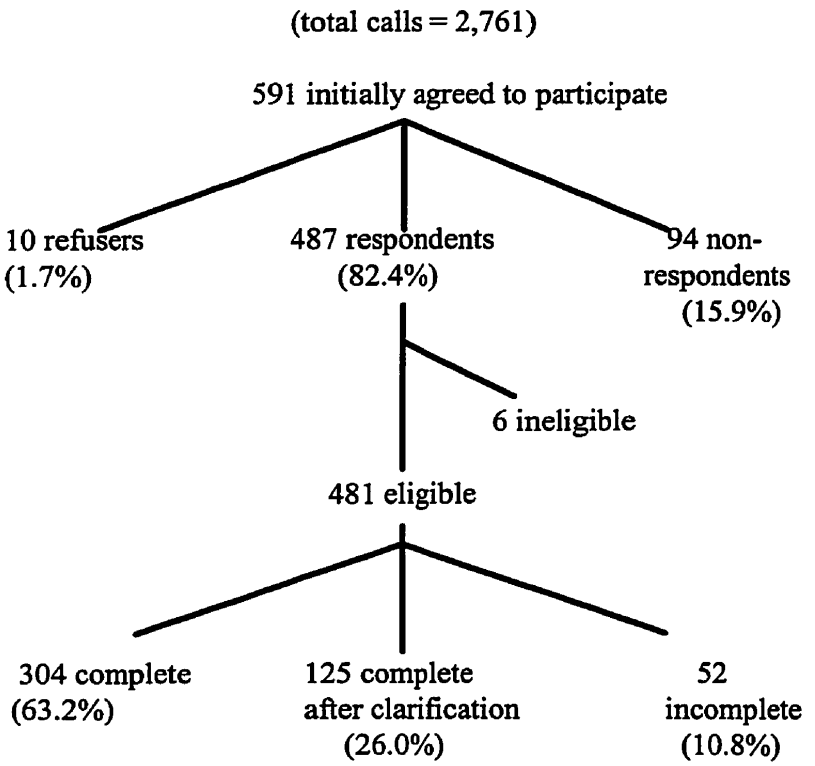
questionnaires were returned 9-11 days after being sent. A second reminder call was placed 10 days after the first reminder call. Reminder calls were delayed if people indicated that they were going to be away or specified they would not have time right away. In order to follow-up, the interviewer made at least five attempts to reach the household over a two-week period, at the end of which the interviewer left a message on an answering machine or with another household member, if possible. No further follow-up was made.

Of the 2,761 listings called, 591 people were eligible and agreed to participate. As the calls were completed, age-sex quotas filled at different rates; toward the end of the study, a greater number of individuals were no longer eligible, which resulted in a low (21%) overall yield. Once an individual had been contacted, if his or her age-sex group was still required, approximately half (47%) of the individuals agreed to participate. The outcome of the initial calls was as follows: agreement (21%), refusal (24%), not eligible (31%), number out of service (9%), unable to contact (9%), comprehension difficulties (3%), and other, such as fax or business, (3%). The goal of recruiting 94 subjects for each age-sex stratum was attained for all strata except males aged 60-75 years, for which only 75 subjects were recruited.

Four hundred and eighty-seven people returned the questionnaire, yielding a return rate of 82.4% (see Figure 3). Six questionnaires were excluded from the analysis: three people were not living in Ontario and three were ineligible due to age.

Three hundred and four questionnaires (63.2%) were complete for MTX intake information. Telephone calls were made to subjects to request clarification or additional information for ***any*** missing MTX information; permission for follow-up was given in 82% of subjects with incomplete information (see Q25, in Appendix B). One hundred and twenty-five questionnaires (26.0%) were completed with a clarification call. The remaining 52 (10.8%) questionnaires remained incomplete, mainly because the subject did not give permission for follow-up. Most missing information was relatively minor. By far the most common missing information pertained to the age at which the individual started drinking chocolate milk or hot chocolate. Few of the remaining incomplete questionnaires contained missing information on the numbers of cups of regular coffee (n=3), tea (n=6) or soft drinks (n=7) consumed per day.

Figure 3. Data collection tree



3.3 Questionnaire

This project used a questionnaire (see Appendix B) designed for a pilot study which examined the association of methylxanthine intake and risk of pancreatic cancer. The six-page questionnaire included demographic and anthropometric data, a selected medical history, exposure to tobacco, a brief dietary history, reproductive history (women only)

and MTX consumption in foods, beverages and medications. The entire questionnaire was estimated to take 20 minutes to complete.

The list of beverages containing MTX included different types of coffee (regular, instant, decaffeinated, espresso, cappuccino), regular tea, chocolate drinks, and cola soft drinks. Subjects were asked to report the age they began drinking a beverage, the usual number of servings per day, and, if applicable, the age at which they stopped.

Subjects were asked to indicate from a list of medications containing MTX, which ones they had used for more than 15 days in any month, or three times a week for three months or more. The breathing medications containing theophylline were as follows: PMS theophylline, Quibron, Slo-bid, Somophyllin, Theo-dur, Theocron, Theolair, Theophylline by injection and Uniphyl. A list of 23 medications containing caffeine was provided. The list contained analgesics, migraine medication and menstrual medication, including Anacin, Excedrin, Fiorinal, Midol, 222, and Tylenol with codeine. Subjects were asked to indicate the usual numbers of pills or injections per day, years used, and usual number of months of use per year, for each medication.

3.4 Data coding and management

The data entry form of the questionnaire was set up in SAS®. A random ten percent of the questionnaires were re-entered into a separate data set and compared to the original data set to evaluate the accuracy of data entry. There was an overall agreement of 99.5%. Data were subsequently cleaned by using simple calculations to check inconsistency (e.g., the reported age the subject stopped drinking a beverage was greater than their current age). Values that were implausible or rare (e.g., a subject who started drinking coffee at age 3 or who reported drinking 40 cups of tea per day) were verified with the original questionnaire response.

When a range of values was given in place of a single value, the median integer value (truncated) was assigned. For example, a typical serving of 4-7 cups of coffee per day was coded as 5. Any consumption less than one unit per day, such as "occasionally", "rarely", or "sometimes" was assigned a value of .5, the median value between 0 and 1. Missing values were coded as 99. A numeric value was assigned to a descriptive age: childhood was coded as 7 years old, and teenager was coded as 15 years old.

Table 1 defines some of the terminology used in the study.

Table 1. Glossary of terms.

Regular coffee	Caffeinated coffee that was percolated or brewed
Total coffee	All types of coffee combined
Total caffeine	Sum of all individual sources of caffeine
Total MTX	Sum of all constituents of MTX
Daily intake (in mg)	(Number of *current* servings per day) x (average caffeine/mtx content per serving)
Lifetime intake (in mg-yrs)	(Number of servings per day) x (average caffeine/mtx content per serving) x (number of years of intake)

Total daily intake of each MTX (caffeine, theobromine, theophylline) was calculated by multiplying servings/dose of each source by its median MTX content (see Tables 2-4 for MTX content used), and then summed over all sources. The tables are based on values adapted from Bullogh *et al.* (1990) and Chou (1992), with additions from the Compendium of Pharmaceuticals and Specialities (1995), the Compendium of Nonprescription Drugs (1995), Ensminger *et al.*, (1994), Stavric *et al.*, (1988) and Watson (1988). The total daily intake of all MTXs was then calculated by summing over the three MTXs. Estimates of lifetime consumption (i.e., analogous to pack-years for smoking) were constructed to account for both quantity and duration of use. Estimates of consumption of each measure were then calculated for each of the six age-sex specific strata.

Table 2. MTX content values (in mg per serving) used for foods.

	Caffeine	Theobromine	Theophylline
Regular coffee	115		
Instant coffee	70		
Decaf. coffee	3		
Espresso, cappuccino	100		
Regular tea	50	6	8
Cocoa, hot chocolate	1	32	
Chocolate milk	6	32	
Cola soft drinks	50		
Chocolate bar	6	10	

Table 3. Caffeine content values (in mg per tablet) used for medications*

	Caffeine
Anacin	32
Atasol	0
Atasol 8, 15, 30	15
Acet. w/codeine	15
C2	15
Cafergot	100
Darvon N	30
Excedrin	65
Fiorinal	40
Midol regular	32.4
Midol PMS	0
Midol extra-strength	60
Novo-Propoxyn	20
217, 222, 282, 292	15
Tylenol with codeine	15

* None of these medications contain theobromine or theophylline.

Table 4. Theophylline values (in mg per tabletl/injection) used for medications.*

	Theophylline
Theophylline by injec.	400
Uniphyl	500
Theodur	400

* None of these medications contain caffeine or theobromine.

Smoking variables were summarized as one categorical variable. Respondents were classified as having never smoked if they indicated that they never used filter cigarettes, non-filter cigarettes or any other tobacco products. Respondents who indicated use of one or more tobacco products at one point, but no current use, were classified as former smokers. Current smokers were respondents who indicated they currently used any tobacco products.

An alcohol variable was created to summarize data reported on the respondents' drinking habits for beer, liquor/spirits, red wine and white wine. A respondent was coded to consume no alcohol, if all four variables were reported to be zero. Consumption was coded as less than one drink per day, if all the alcohol variables were less than 1-6 per week **or** if less than three out of the four variables were reported to be 1-6 drinks per week; otherwise, consumption was coded to be one or more drinks per day.

3.5 Statistical analyses

Descriptive analyses were used to present subject characteristics and MTX data. The number of servings of MTX-containing beverages were presented by age and sex, as well as the mean values, standard variations and maximum values for daily and lifetime use for each MTX-containing substance (in mg). Some comparisons of the MTX intake were made by region, education level, smoking status and alcohol use.

Three methods were used to assess the misclassification caused by using coffee as an indicator of caffeine and MTX intake. First, the percentage of caffeine intake from regular coffee as a total caffeine intake and the percentage of MTX intake from regular coffee as a total of MTX intake were estimated. Second, caffeine intake (in mg) was categorized into four levels of exposure based on quartiles of caffeine intake from regular coffee consumption. Four categories were used because this is the number of exposure levels generally reported in the literature. Cut-points were chosen to allow, as much as possible, an even distribution of coffee exposure among categories. Cross-tabulations were constructed for categorized versions of each continuous variable, showing the classification that would result when using coffee-only versus caffeine, and coffee-only versus MTX (see Figure 4). Calculations were performed on estimates of present and lifetime use by age and sex.

Figure 4. Classification matrix: caffeine intake from regular coffee only (in mg) by caffeine intake from all sources (in mg)*

Level of exposure Caffeine intake

Coffee intake

	E_1	E_2	E_3	E_4	
E_1	$n_{1,1}$	$n_{1,2}$	$n_{1,3}$	$n_{1,4}$	$n_{1,1-4}$
E_2	0**	$n_{2,2}$	$n_{2,3}$	$n_{2,4}$	$n_{2,1-4}$
E_3	0	0	$n_{3,3}$	$n_{3,4}$	$n_{3,1-4}$
E_4	0	0	0	$n_{4,4}$	$n_{4,1-4}$

* $n_{1,1}$ is the number of subjects in the sample who reported the lowest exposure level regardless of whether caffeine intake was being estimated by coffee only or caffeine from all sources; $n_{1,2}$ is the number of subjects who reported the lowest exposure level estimated by coffee only, but who were in the second exposure level estimated by total caffeine intake; etc.
** Cells below the main diagonal will be equal to zero because the coffee only condition will always lead to total caffeine intake that is equal to or greater than that of coffee.

The classification matrix provides the percentages of respondents which were correctly classified from coffee, along with the percentages which should have been reported in each of the higher levels of exposure.

Finally, the kappa statistic was used to assess the level of agreement between classification that would result from using coffee-only versus caffeine, and coffee-only versus MTXs. Weighted kappa statistics were used for multiple levels of exposure using squared error weights (Fleiss, 1981). Six categories were used, this number was chosen arbitrarily because of the lack of clear guidelines on this matter in the literature (Brenner & Kliebsch, 1996). Formulae and further explanations are provided in Appendix D.

The effect of misclassification on odds ratio estimates was assessed by two methods. The first method set hypothetical odds ratios and estimated the misclassified odds ratios resulting from estimates of caffeine intake from coffee only. The second method involved correcting published odds ratio estimates. Both methods of analysis are presented in Kleinbaum *et al.* (1982) for two levels of exposure. This study extended the two levels to apply to multiple levels of exposure. Starting and derived formulae for the first method are presented in the next section. Briefly, four categories of exposure were defined by intake from regular coffee reported in this sample. The number of "controls" in each exposure level was defined by the levels of total caffeine intake from coffee reported among this sample. A hypothetical case distribution was constructed to provide various odds ratio estimates reflecting a threshold and dose-response relationship. Subjects were then reclassified into the caffeine exposure level that would have occurred had they been estimated by coffee only, using results from Figure 4 and derived formulae, assuming non-differential misclassification for cases and controls. The resulting odds ratio estimates were compared to the true odds ratio estimates.

The second method involved applying corrections to the crude odds ratio estimates in studies in the literature which used coffee as a measure. The MacMahon *et al.*, (1981) study was chosen for its historical importance. The limitation of using this study is that

caffeine patterns are likely to have changed in the last 15 years and American consumption patterns may be different from Canadian patterns. Thus, a second article (Jain *et al.*, 1991) was chosen; it was a fairly recent study conducted in Canada and therefore more likely to be similar to the present study sample. The corrections to the data in these papers provide a way of examining the type of effect that might be expected ***if*** an assumption was made that coffee is a good surrogate measure for caffeine. See Appendix E for formulae.

3.6 Formulae

Formulae to estimate the effect of underestimation on the OR for two levels of exposure are presented. These simpler formulae help clarify the method used. They are then extended to apply to multiple levels of exposure. Finally, a sample calculation is provided to help illustrate the multiple level of exposure situation which is used in subsequent analyses.

3.6.1 Two levels of exposure

These formulae are taken from Kleinbaum *et al.* (1982), pp. 220-232, given non-differential misclassification for disease and exposure. Specificity for this study is 1.0 since estimates of caffeine intake from coffee will never result in an underestimation of total caffeine intake. Formulae from Kleinbaum *et al.*(1982) were simplified to meet this condition.

Let a 2x2 true classification of caffeine intake from all sources be defined as:

	E_{T1}	E_{T0}
D_{T1}	A_T	B_T
D_{T0}	C_T	D_T

(1)

where E_{T1} represents exposure to caffeine, E_{T0} non-exposure, D_{T1} cases, and D_{T0} controls. Then, A_T is the hypothesized number of cases exposed to caffeine, B_T is the hypothesized number of cases not exposed to caffeine, C_T is the number of controls in this study who reported being exposed to caffeine, and D_T is the number of controls in this study who reported not being exposed to caffeine.

For simplicity the number of cases and controls was set to equal 100.

Let the odds ratio estimate be defined as, $OR_T = \frac{A_T * D_T}{B_T * C_T}$. (2)

The OR was set to the desired value. By simple algebra it can be shown from (1) and (2), where $A_T + B_T = 100$:

	E_{T1}	E_{T0}
D_{T1}	$\frac{100 * C_T * OR}{C_T * OR + D_T}$	$100 - A_T$
D_{T0}	C_T	D_T

(3)

where values for C_T and D_T were based on exposure estimates from this study.

Let us define a classification matrix of caffeine intake from regular coffee intake only (in mg) by caffeine intake from all sources (in mg), where values of $n_{0,0}$, $n_{0,1}$ and $n_{1,1}$ were dependent on the data collected in this study.

	True caffeine exposure	
Classified exp.	E_0	E_1
E_0	$n_{0,0}$	$n_{0,1}$
E_1	0*	$n_{1,1}$

(4)

*$n_{1,0}$ will always equal 0, because estimates of caffeine intake from coffee only will always underestimate total caffeine intake.

ϕ_E represents the sensitivity, that is, the probability that a person who is exposed will be classified as exposed or $\phi_E = n_{1,1} / (n_{0,1} + n_{1,1})$. (5)

Let a 2x2 actual population of caffeine intake using regular coffee only be defined as:

	E_{M1}	E_{M0}
D_{M1}	A_M	B_M
D_{M0}	C_M	D_M

(6)

where E_{M1} represents exposure to caffeine from coffee, E_{M0} non-exposure, D_{M1} cases, and D_{M0} controls. Then, A_M is the hypothesized number of cases exposed to coffee, B_M is the hypothesized number of cases not exposed to coffee, C_M is the number of controls in this study who reported being exposed to coffee, and D_M is the number of controls in this study who reported not being exposed to coffee.

Given that the specificity in this study is 1.0 and assuming non-differential misclassification of disease, the formulae provided can be simplified to show that:

$$\begin{aligned} A_M &= A_T * \phi_E \\ B_M &= (1-\phi_E) * A_T + B_T \\ C_M &= C_T * \phi_E \\ D_M &= (1-\phi_E) * C_T + D_T \end{aligned} \qquad (7)$$

$$OR_M = \frac{A_M * D_M}{B_M * C_M}. \qquad (8)$$

To estimate the effect of misclassification on the odds ratio, OR_T (2) is compared to OR_M (8).

3.6.2 Multiple levels of exposure

The same formulae can be extended for multiple levels of exposure. In four levels of exposure, exposure categories are collapsed to form 2x2 tables in three possible combinations. That is, 1) the lowest level of exposure is preserved while the three higher levels are collapsed into one, 2) the first three levels of exposure are collapsed into one while the highest level of exposure is preserved, and 3) the lowest two are collapsed to form one level, as are the two highest levels.

Let a 4-level exposure for true classification of caffeine intake from all sources be defined as:

	E_{T1}	E_{T2}	E_{T3}	E_{T4}
D_{T1}	A_T	B_T	C_T	D_T
D_{T0}	E_T	F_T	G_T	H_T

(9)

where E_{T1} represents the lowest level of exposure, E_{T2} the second level of exposure, E_{T3} the third level of exposure, E_{T4} the highest level of exposure, D_{T1} the cases, and D_{T0} the controls.

Let the number of cases and controls equal 100 such that:

$$A_T + B_T + C_T + D_T = 100$$
$$E_T + F_T + G_T + H_T = 100. \qquad (10)$$

Let the OR estimates be defined as:

$$OR_{T1,1} = 1.0$$
$$OR_{T2,1} = (B_T * E_T)/(A_T * F_T)$$
$$OR_{T3,1} = (C_T * E_T)/(A_T * G_T) \qquad (11)$$
$$OR_{T4,1} = (D_T * E_T)/(A_T * H_T).$$

The $OR_{T1,1}$, $OR_{T2,1}$, $OR_{T3,1}$, $OR_{T4,1}$ are set at desired values. Then by simple algebra it can be shown from (10) and (11) that:

	E_{T1}	E_{T2}	E_{T3}	E_{T4}
D_{T1}	$\frac{100*E_T}{(E_T+F_T*OR_{T2,1}+G_T*OR_{T3,1}+H_T*OR_{T4,1})}$	$\frac{A_T*F_T*OR_{T2,1}}{E_T}$	$\frac{A_T*G_T*OR_{T3,1}}{E_T}$	$\frac{A_T*H_T*OR_{T4,1}}{E_T}$
D_{T0}	E_T	F_T	G_T	H_T

(12)

Let the classification matrix for caffeine intake from regular coffee intake only (in mg) by caffeine intake from all sources (in mg)* be defined as:

Caffeine (13)

Coffee	E_1	E_2	E_3	E_4	
E_1	$n_{1,1}$	$n_{1,2}$	$n_{1,3}$	$n_{1,4}$	$n_{1,1-4}$
E_2	0*	$n_{2,2}$	$n_{2,3}$	$n_{2,4}$	$n_{2,1-4}$
E_3	0	0	$n_{3,3}$	$n_{3,4}$	$n_{3,1-4}$
E_4	0	0	0	$n_{4,4}$	$n_{4,1-4}$

* Cells below the main diagonal will be equal to zero because the coffee only condition will always lead to total caffeine intake that is greater than or equal to that of coffee.

Let a 4-level classification for actual population intake of caffeine using regular coffee only be:

	E_{M1}	E_{M2}	E_{M3}	E_{M4}
D_{M1}	A_M	B_M	C_M	D_M
D_{M0}	E_M	F_M	G_M	H_M

(14)

where E_{M1}, E_{M2}, E_{M3} and E_{M4} represent the level of exposure to caffeine from coffee, D_{M1} represents cases and D_{M0} controls. Then, A_M, B_M, C_M, and D_M represent the estimated level of exposure of caffeine from coffee for cases, and E_M, F_M, G_M and H_M the estimated level for controls.

Then, the odds ratios are defined as:

$$\begin{aligned} OR_{M1,1} &= 1.0 \\ OR_{M2,1} &= (B_M * E_M)/(A_M * F_M) \\ OR_{M3,1} &= (C_M * E_M)/(A_M * G_M) \\ OR_{M4,1} &= (D_M * E_M)/(A_M * H_M). \end{aligned} \qquad (15)$$

The four levels of exposures can be collapsed into 2x2 tables in three possible ways. The formulae then translate from (5) and (7) in the following way:

Step 1: Collapse exposure levels 2 through 4.

	E_{T2-4}	E_{T1}
D_{T1}	$B_T+C_T+D_T$	A_T
D_{T0}	$F_T+G_T+H_T$	E_T

$$\phi_{E1,\,E2-4} = (n_{2,2-4} + n_{3,2-4} + n_{4,2-4}) \,/\, (n_{1,2-4} + n_{2,2-4} + n_{3,2-4} + n_{4,2-4})$$

$B_M+C_M+D_M = (B_T+C_T+D_T) * \phi_{E1, E2-4}$
$A_M = (1- \phi_{E1, E2-4})* (B_T+C_T+D_T) + A_T$ (16)
$F_M+G_M+H_M = (F_T+G_T+H_T) * \phi_{E1, E2-4}$
$E_M = (1- \phi_{E1, E2-4})*(F_T+G_T+H_T) + E_T$

Step 2: Collapse exposure levels 1 through 3.

	E_{T4}	E_{T1-3}
D_{T1}	D_T	$A_T+B_T+C_T$
D_{T0}	H_T	$E_T+F_T+G_T$

$\phi_{E1-3, E4} = n_{4,4} / n_{1-4,4}$

$D_M = D_T* \phi_{E1-3, E4}$ (17)
$A_M+B_M+C_M = (1- \phi_{E1-3, E4}) * D_T + (A_T+B_T+C_T)$
$H_M = H_T * \phi_{E1-3, E4}$
$E_M+F_M+G_M = (1- \phi_{E1-3, E4}) * H_T + (E_T+F_T+G_T)$

Step 3: Collapse exposure levels 1 and 2, and levels 3 and 4.

	E_{T3-4}	E_{T1-2}
D_{T1}	C_T+D_T	A_T+B_T
D_{T0}	G_T+H_T	E_T+F_T

$\phi_{E1-2, E3-4} = (n_{3,3-4} + n_{4,3-4}) / (n_{1,3-4} + n_{2,3-4} + n_{3,3-4} + n_{4,3-4})$

$C_M+D_M = (C_T+D_T) * \phi_{E1-2,E3-4}$ (18)
$A_M+B_M = (1- \phi_{E1-2,E3-4}) * (C_T+D_T) + (A_T+B_T)$
$G_M+H_M = (G_T+H_T) * \phi_{E1-2,E3-4}$
$E_M+F_M = (1-\phi_{E1-2,E3-4}) *(G_T+H_T) + (E_T+F_T)$

Solve for A_M, B_M, C_M and D_M (16)(17)(18). Values of A_M, B_M, C_M and D_M are then substituted in (14) and OR_M can be calculated from (15). To estimate the effect of misclassification bias on the odds ratio estimates, the true OR_T (11) were compared to the misclassified OR_M (15).

3.6.3 Working example

Numbers for this example ***do not reflect values found in this study.*** They were chosen for simplicity and are used to illustrate the methods. Hypothetical values are highlighted in ***bold.***

Let a 4-level exposure for true classification of caffeine intake from all sources be defined as:

	E_{T1}	E_{T2}	E_{T3}	E_{T4}
D_{T1}	10	15	30	45
D_{T0}	**20**	**20**	**30**	**30**

Then,

$OR_{T1,1} = 1.0$
$OR_{T2,1} = (15*20)/(10*20) = \mathbf{1.5}$
$OR_{T3,1} = (30*20)/(10*30) = \mathbf{2.0}$
$OR_{T4,1} = (45*20)/(10*30) = \mathbf{3.0}$.

Let the classification matrix for caffeine intake from regular coffee intake only (in mg) by caffeine intake from all sources (in mg) be defined as:

	Caffeine				
Coffee	E_1	E_2	E_3	E_4	
E_1	**40**	**30**	**20**	**10**	100
E_2	0	**70**	**60**	**50**	180
E_3	0	0	**90**	**80**	170
E_4	0	0	0	**100**	100
					550

Let a 4-level classification for actual population intake of caffeine using regular coffee only be defined as:

	E_{M1}	E_{M2}	E_{M3}	E_{M4}
D_{M1}	A_M	B_M	C_M	D_M
D_{M0}	E_M	F_M	G_M	H_M

In order to obtain values for A_M-H_M the four levels of exposure are collapsed into 2x2 tables in three possible ways.

Step 1: Collapse exposure levels 2 through 4.

	E_{T2-4}	E_{T1}
D_{T1}	90	10
D_{T0}	80	20

$\phi_{E1, E2-4} = (70+60+50+90+80+100)/(30+20+10+70+60+50+90+80+100) =$
$450/510 = 0.8824$
$A_M = (1-0.8824)* 90 + 10 = 20.6$
$E_M = (1-.8824)*80 + 20 = 29.4$

Step 2: Collapse exposure levels 1 though 3.

	E_{T4}	E_{T1-3}
D_{T1}	45	55
D_{T0}	30	60

$\phi_{E1-3, E4} = 100/(10+50+80+100) = 100/240 = 0.4167$
$D_M = 45* 0.4167 = 18.8$
$H_M = 30 * 0.4167 = 12.5$

Step 3: Collapse exposure levels 1 and 2, and levels 3 and 4.

	E_{3-4}	E_{1-2}
D	75	25
~D	60	40

$\phi_{E1-2, E3-4} = (90+80+100)/(20+10+60+50+90+80+100) =$
$270/410 = 0.6585$

$C_M+D_M = 75 * \phi_{E1-2,E3-4} <=> C_M = 75 * \phi_{E1-2,E3-4} - D_M =$
$75 * 0.6585 - 18.8 = 30.6$

$A_M+B_M = (1- \phi_{E1-2,E3-4}) * (C+D) + (A+B) <=>$
$B_M = (1- \phi_{E1-2,E3-4}) * (C+D) + (A+B) - A_M = (1-0.6585)*75+25 - 20.6 = 30.0$

$G_M+H_M = (G+H) * \phi_{E1-2,E3-4} <=> G_M = (G+H) * \phi_{E1-2,E3-4} - H_M =$
$60*0.6585 - 12.5 = 27.0$

$E_M+F_M = (1-\phi_{E1-2,E3-4}) *(G+H) + (E+F) <=>$
$F_M = (1-\phi_{E1-2,E3-4}) *(G+H) + (E+F) - E_M = (1-0.6585)*60 +40 - 29.4 = 31.1$

Therefore the cells for 4-level classification for actual population intake of caffeine from regular coffee can be filled in:

	E_{M1}	E_{M2}	E_{M3}	E_{M4}
D_{M1}	20.6	30.0	30.6	18.8
D_{M0}	29.4	31.1	27.0	12.5

The misclassified odds ratio estimates are:
$OR_{M1,1} = 1.0$
$OR_{M2,1} = (30.0*29.4)/(20.6*31.1) = 1.4$
$OR_{M3,1} = (30.6*29.4)/(20.6*27.0) = 1.6$
$OR_{M4,1} = (18.8*29.4)/(20.6*12.5) = 2.1$

In this hypothetical example, true odds ratio estimates of 1.0, 1.5, 2.0, 3.0 would have been misclassified as 1.0, 1.4, 1.6, 2.1. Effects on the preselected odds ratio depend on values in the classification matrix as well as the distribution of controls among exposure levels. The values used in ***subsequent*** analyses were based on reports of coffee and total caffeine intake from the subjects in this study.

4. RESULTS

This section is divided into three main sections, each covering one study objective. While each objective provides important information on its own, each individual objective builds on the others. Findings on coffee, caffeine and MTX intake were used to build underestimates of using coffee as an index of caffeine intake. These estimates were used in turn to measure the effects of underestimation on the odds ratio.

4.1 Objective 1: Estimate MTX intake

This subsection gives a brief description of the characteristics of the sample used in this study, describes MTX intake by presenting serving sizes of MTX-containing beverages, and provides summary results for users by beverage. It presents preliminary results on the relationship of coffee to total caffeine intake and the mean daily and lifetime intake from different sources of caffeine and MTX. It also presents the data by region and by known associated factors: tobacco use, education, and alcohol use.

The distribution of the 481 respondents by age and sex is shown in Table 5.

Table 5. Frequency distribution (number, percent) of respondents by age and sex.

	30 - 44 years	45-59 years	60-75 years	Total
Male	85 (17.7%)	73 (15.1%)	59 (12.3%)	217 (45.1%)
Female	80 (16.7%)	90 (18.7%)	94 (19.5%)	264 (54.9%)

The distribution of education level, smoking status, and alcohol intake for each age-sex stratum are reported in Table 6. These factors are important to describe because they have been reported to be potential confounders for coffee intake. For both males and females, the reported level of formal education decreased as age increased. Smoking status varied by age and sex; generally, the younger individuals were more likely to report daily tobacco use. Also, males reported a higher consumption of alcohol than did females.

Table 6: Percent distribution of respondent characteristics by age (in years) and sex.

	Male			Female		
	30-44 years n=85	45-59 years n=73	60-75 years n=59	30-44 years n=80	45-59 years n=90	60-75 years n=94
Education			(2)		(1)	
< Grade 9	2.4	5.5	22.0	0.0	11.1	16.0
Grade 9-12	20.0	43.8	35.6	23.8	32.2	47.9
Uni/College	77.6	50.7	39.0	76.3	55.6	36.2
Smoking status						(1)
Never	55.3	20.5	30.5	47.5	50.0	54.2
Former	17.6	52.1	55.9	30.0	31.1	32.9
Daily	27.1	27.4	13.6	22.5	18.9	11.7
Alcohol intake						
Never	14.1	16.4	18.6	23.8	17.8	27.7
Occasional	56.5	52.1	44.1	66.3	64.4	60.6
Daily	29.4	31.5	37.3	10.0	17.8	11.7

The numbers in parentheses indicate the number of unknown values for that specific characteristic.

4.1.2 Current servings of MTX-containing beverages

Regular coffee was the most common type of coffee consumed (see Table 7). Overall, three quarters of the subjects reported drinking regular coffee daily; less than a quarter reported drinking instant coffee, decaffeinated coffee or espresso/cappuccino on a daily basis. Patterns of coffee intake varied by age and sex; females aged 30-44 and females aged 60-75 were the least likely to report a daily use of caffeinated coffees (regular

were the most likely to report daily intake of caffeinated coffee (see Table 7).

Table 7. Percent distribution of current servings per day of different types of coffee beverages among respondents by age and sex.

	Male			Female		
	30-44 n=85	45-59 n=73	60-75 n=59	30-44 n=80	45-59 n=90	60-75 n=94
Regular coffee		(1)*				(2)
Non-drinker	22.9	29.2	25.9	38.5	32.2	40.7
<1	2.4	2.8	3.4	2.6	7.8	5.5
1	24.1	15.3	12.1	24.4	18.9	19.8
2	31.7	15.3	24.1	15.4	26.7	18.7
≥3	16.9	37.6	34.4	25.9	14.4	15.4
Instant coffee		(1)		(1)		
Non-drinker	79.5	68.1	74.1	75.6	74.4	63.7
<1	6.0	4.2	3.4	0.0	4.4	6.6
1	7.2	8.3	3.4	16.7	14.4	15.4
2	1.2	8.3	10.3	5.1	4.4	5.5
≥3	6.0	11.1	8.6	2.6	2.2	8.8
Decaf. coffee						
Non-drinker	88.2	86.3	78.0	82.5	78.9	74.5
<1	3.5	4.1	6.8	1.3	3.3	1.6
1	3.5	2.7	1.7	12.5	10.0	5.3
2	2.4	2.7	10.2	3.8	5.6	4.3
≥3	2.4	4.1	3.4	0.0	2.2	5.3
Espresso/cappuc.						
Non-drinker	78.8	93.2	93.2	83.8	82.2	95.7
<1	12.9	5.1	5.1	10.0	11.1	3.2
1	8.2	0.0	0.0	6.3	2.2	1.1
≥2	0.0	1.7	1.7	0.0	4.4	0.0
All types combined**			(1)			
<1	11.8	9.6	6.7	25.3	13.3	13.3
1	18.8	13.7	8.5	9.2	16.7	18.1
2	20.0	20.6	20.3	24.1	27.8	27.8
3	22.3	17.8	30.5	21.5	17.7	17.7
≥4	27.1	38.3	34.0	18.9	24.5	23.1

* The number in parentheses indicates the number of unknown values for that specific category.
** This variable was derived by summing the number of servings from each type of coffee.

highest percentage of daily tea drinkers (68.2%) was reported among females aged 60-75; the lowest intake (26.5%) was among males aged 30-44.

Cola soft drink intake varied across age and sex groups. For both males and females, the proportion which reported drinking cola soft drinks decreased by age (see Table 8). One example illustrating the difference of drinking patterns between coffee and cola drinkers is that the vast majority of coffee drinkers drank coffee on a daily basis, while a much higher percentage of cola drinkers drank cola on only an occasional basis.

Hot chocolate and chocolate milk were less popular than regular coffee and regular tea (see Table 8). A small proportion of subjects reported daily intake and, of those, the number of servings per day was low.

beverages, excluding coffee, among respondents by age and sex.

	Male			Female		
	30-44 n=85	45-59 n=73	60-75 n=59	30-44 n=80	45-59 n=90	60-75 n=94
Regular tea		(4)*		(1)		(1)
Non-drinker	54.2	47.2	37.9	24.4	23.3	17.6
<1	19.3	9.7	3.4	16.7	16.7	14.3
1	16.9	15.3	20.7	29.5	23.3	28.6
2	1.2	18.1	19.0	11.5	14.4	16.5
≥3	8.4	9.7	19.0	17.9	22.2	23.1
Cola drinks		(4)		(2)		(1)
Non-drinker	16.9	34.7	39.7	32.1	40.0	63.7
<1	36.1	22.2	34.5	29.5	25.6	24.2
1	27.7	16.7	15.5	23.1	24.4	6.6
2	14.5	15.3	5.2	10.3	5.6	3.3
≥3	4.8	11.1	5.2	5.1	4.4	2.2
Hot chocolate						
Non-drinker	69.4	89.0	81.4	68.8	77.8	69.1
<1	28.2	8.2	18.6	25.0	17.8	24.5
1	0.0	0.0	0.0	6.3	4.4	4.3
≥2	0.0	0.0	0.0	0.0	0.0	2.2
Chocolate milk						
Non-drinker	70.6	87.7	88.1	77.5	84.4	84.0
<1	23.5	11.0	11.9	17.5	10.0	12.8
1	3.5	1.4	0.0	5.0	4.4	2.1
≥2	2.4	0.0	0.0	0.0	1.1	1.1

*The number in parentheses indicates the number of unknown values for that specific category.

The average serving intake for users of MTX-containing beverages is displayed in Tables 9 and 10. The mean was reported instead of the median, because mean values allowed comparison with other data (e.g., Ontario Health Survey, Coffee Association of Canada). Regular coffee, instant coffee, decaffeinated coffee and regular tea had the highest daily serving size ranging from 1.5 - 2.7 cups per day. The remaining beverages all had a mean intake of approximately 1 cup per day or slightly lower.

beverages for **male users*** by age.

	30-44 years			45-59 years			60-75 years		
	mean	s.d.	max	mean	s.d.	max	mean	s.d.	max
Regular coffee	2.3	1.7	12	3.0	2.8	20	2.8	2.7	20
Instant coffee	1.5	1.2	5	1.9	1.3	5	2.0	1.3	5
Decaf. coffee	1.4	0.9	3	1.5	1.2	4	1.8	1.4	5
Espresso/cappuc.	0.9	1.0	5	0.8	0.5	2	0.8	1.3	2
Regular tea	1.4	1.2	5	1.8	1.5	6	2.2	1.5	6
Cola drinks	1.2	1.2	8	1.6	1.6	9	1.1	1.1	6
Hot chocolate	0.6	0.3	2	0.6	0.2	1	0.6	0.2	1
Chocolate milk	0.7	0.4	2	1.2	2.1	8	0.6	0.2	1

* Occasional and daily users were included in these calculations.

Table 10. Mean, standard deviation and maximum of servings of MTX-containing beverages for **female users*** by age.

	30-44 years			45-59 years			60-75 years		
	mean	s.d.	max	mean	s.d.	max	mean	s.d.	max
Regular coffee	2.0	1.5	8	2.2	1.8	12	2.0	1.5	8
Instant coffee	1.6	1.5	8	1.7	1.5	8	2.0	1.9	8
Espresso/cappuc.	0.7	0.2	1	0.9	0.6	2	1.0	1.0	3
Decaf. coffee	1.9	2.7	12	1.6	1.2	12	2.0	1.8	7
Cola drinks	1.1	0.9	5	1.1	0.8	5	0.9	0.8	4
Regular tea	1.7	1.7	10	2.0	1.4	6	2.7	4.6	40
Hot chocolate	0.7	0.5	3	0.6	0.2	1	0.9	0.9	5
Chocolate milk	0.9	0.6	3	0.8	0.5	2	0.8	0.9	5

* Occasional and daily users were included in these calculations.

The following subsections report the mean intake of caffeine and MTX intake in milligrams, rather than serving size, in order to highlight the contribution of each MTX-containing substance to total caffeine or MTX intake. Once again, mean values are used for purposes of comparison with other data sources.

In all strata, the four main sources of caffeine were regular coffee, regular tea, cola drinks and instant coffee (see Table 11 and Figure 5). Combined, these four sources accounted for 90.1% to 98.3% of the total daily caffeine intake.

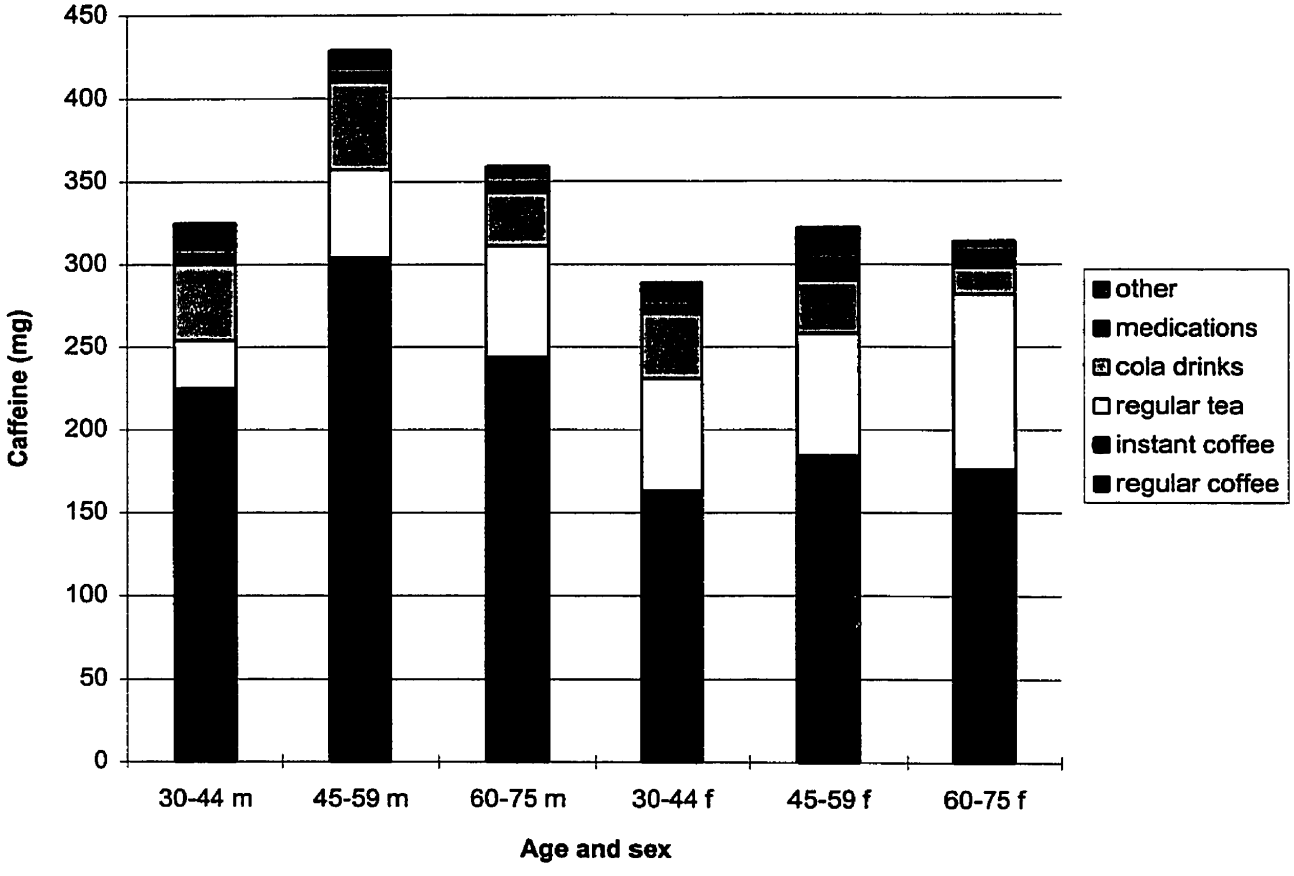
Coffee was the main source of caffeine intake for all strata (see Table 11 and Figure 5). Regular tea was the second highest source of caffeine intake, except for males aged 30-44 where the caffeine contribution from cola soft drinks was higher. In females aged 60-75 years, tea was almost as high a source of caffeine as coffee. The daily regular coffee and cola soft drink intake were higher for males than females. In contrast, the daily regular tea intake was lower for males than females. The contribution of caffeine from medications was minimal for all strata.

average percent of total caffeine intake among respondents by age and sex.

	Male						Female					
	30-44 y		45-59 y		60-75 y		30-44 y		45-59 y		60-75 y	
	mg	% caff	mg	% caff	mg	% caff	mg	% caff	mg	% caff	mg	% ca
Regular coffee	204	62.8	259[1]	60.8	205	57.1	139	48.1	159	49.4	131[2]	41.
Instant coffee	21	6.5	45[1]	10.6	39	10.9	24[1]	8.3	25	7.8	45	14.
Regular tea	29	8.9	53[4]	12.4	67	18.7	68[1]	23.5	74	23.0	106[1]	33.
Cola drinks	46	14.1	53[4]	12.4	32	8.9	39[2]	13.5	32	10.0	16[1]	5.
Medications	6	2.0	5	1.2	6	1.7	5	1.7	13	4.0	10	3.
Other (e.g.,choc bars)	19	5.8	11	2.6	10	2.8	14	4.8	19	5.9	6	1.
Total	325	100.1	426	100.0	359	100.1	289	99.9	322	100.1	314	10

Superscripts indicate the number of unknown values.

Figure 5. Mean daily caffeine intake from different sources (in mg) among respondents by age and sex.



MTX daily intake (see Tables 11 and 12). While males aged 30-44 reported drinking the same amount of regular coffee as males aged 60-75, those in the 30-44 year age group had a lower total caffeine and MTX daily intake. Similarly, while females aged 30-44 reported drinking slightly more coffee than females aged 60-75, the younger age group had a lower total daily caffeine and MTX intake.

Because females reported drinking more tea than did males, and because tea contains both theobromine and theophylline, the sex differences for MTX intake are less than the sex difference for caffeine intake. Overall, the average daily intake of caffeine was 333 mg. The average intake of total MTX was 364 mg; theophylline and theobromine combined provided only 31 mg. Caffeine as a source of daily MTX ranged from 87.8% to 95.4%; theobromine ranged from 2.7% to 7.4% of MTX intake; and theophylline ranged from 1.4% to 4.8% (see Table 12).

Table 12. Mean daily intake of different constituents of MTX (in mg) and average percent of total MTX intake among respondents by age and sex.

	Male						Female					
	30-44 y		45-59 y		60-75 y		30-44 y		45-59 y		60-75 y	
	mg	% MTX	mg	% MTX	mg	% MTX	mg	% MTX	mg	% MTX	mg	% MTX
Total caffeine[1]	325	93.3	417	95.4	359	92.5	287	90.3	322	90.4	309	87.8
Total theobromine	18	5.2	12	2.7	14	3.6	20	6.3	18	5.1	26	7.4
Total theophylline	5	1.4	8	1.8	15	3.9	11	3.5	16	5.0	17	4.8
Total methylxanthine	348	99.9	437	99.9	388	100.0	318	100.1	356	100.0	352	100.0

[1]The total caffeine intake indicated here may be slightly lower than summing caffeine intake from different sources (Table 11) because unknown values were set to zero here whereas in the previous table they were dropped from the calculation. The values were set to zero in this case to avoid dropping data due to missing values. It was discovered during follow-up calls that missing values tended to indicate occasional use only.

4.1.4 Lifetime MTX intake

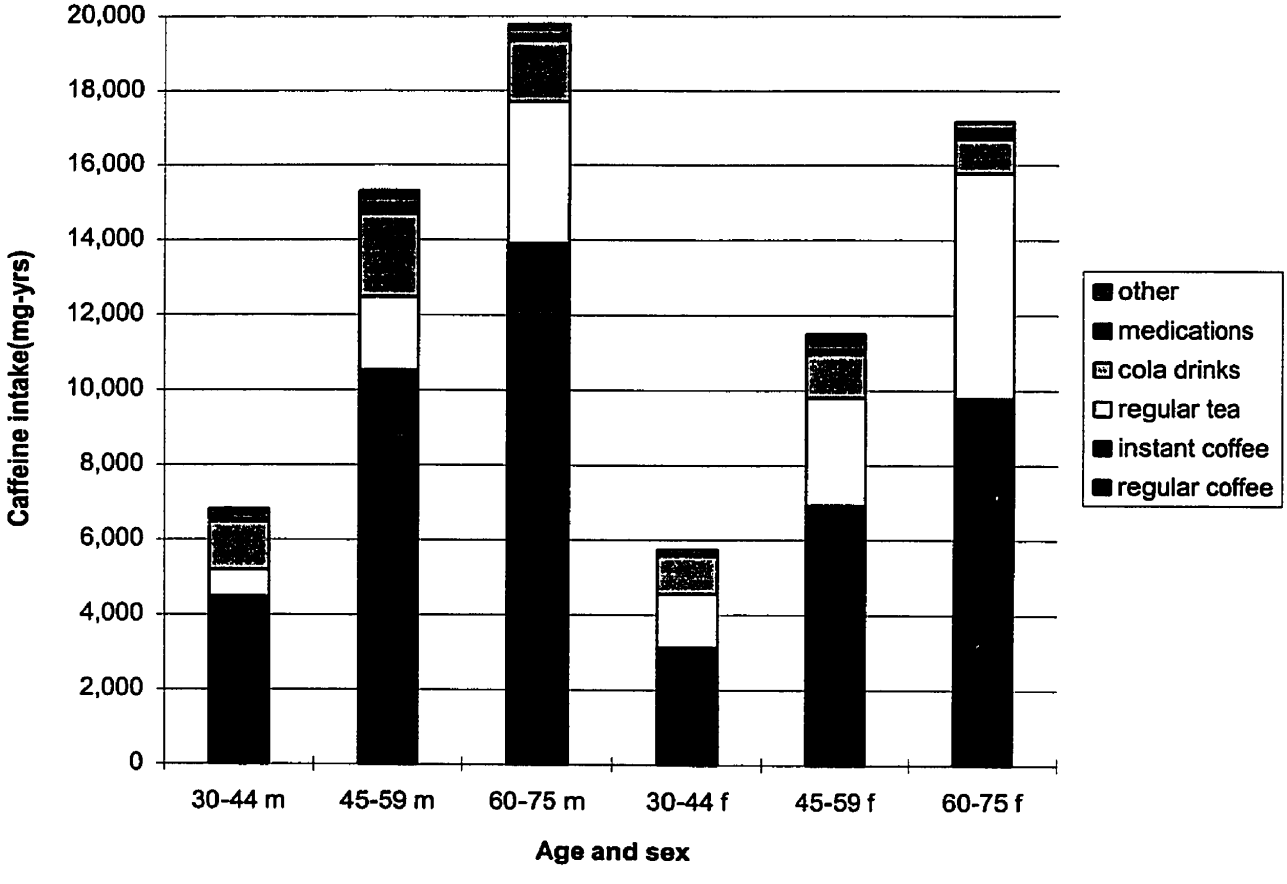
Lifetime values of MTX were expressed in mg-years of intake. Since the amounts of caffeine and MTX are cumulative in this calculation, older subjects are expected to have higher intake. Lifetime patterns of both caffeine and MTX intake followed patterns similar to those of daily intake in terms of the main sources of both caffeine (compare Tables 11 and 13; compare Figures 5 and 6) and MTX (compare Tables 12 and 14).

average percent of total caffeine intake among respondents by age and sex.

	Males						Females					
	30-44 y		45-59 y		60-75 y		30-44 y		45-59 y		60-75 y	
	mg	% caff	mg	% caff	mg	% caff	mg	% caff	mg	% caff	mg	% caff
Regular coffee	3,991	58.1	8,857[3]	57.8	12,143	61.4	2,551[2]	44.3	5,824[2]	50.6	7,789[6]	45.4
Instant coffee	490[2]	7.1	1,674[2]	10.9	1,754[1]	8.9	574[4]	10.0	1,101[2]	9.6	1,998[2]	11.6
Regular tea	708[2]	10.3	1,942[5]	12.7	3,797[1]	19.2	1,438[4]	25.0	2,874[2]	25.0	5,992[4]	34.9
Cola drinks	1,290[1]	18.8	2,210[6]	14.4	1,644	8.3	1,016[4]	17.6	1,159[2]	10.1	917[7]	5.3
Medications	37	0.5	267	1.7	148	0.7	50[1]	0.9	114	1.0	272	1.6
Other (e.g.,choc bars)	350	5.1	366	2.4	389	1.5	164	2.2	439	3.8	203	1.2
Total	6,866	99.9	15,316	99.9	19,875	100.0	5,793	100.0	11,511	100.1	17,171	100.0

The superscripts indicate the number of unknown values.

Figure 6. Mean lifetime intake of caffeine from different sources (in mg-yrs) among respondents by age and sex.



average percent of total MTX intake among respondents by age and sex.

	Male						Female					
	30-44 y		45-59 y		60-75 y		30-44 y		45-59 y		60-75 y	
	mg	% MTX	mg	% MTX	mg	% MTX	mg	% MTX	mg	% MTX	mg	% MTX
Total caffeine[1]	6,822	91.7	14,592	94.7	19,782	93.5	5,575	87.3	11,268	87.9	16,308	87.3
Total theobromine	506	6.8	521	3.4	786	3.7	565	8.8	732	5.8	1,458	7.8
Total theophylline	111	1.5	289	1.9	597	2.8	249	3.9	534	4.3	918	4.9
Total methylxanthine	7,439	100.0	16,127	100.0	21,172	100.0	6,391	100.0	12,534	100.0	18,684	100.0

[1] The total caffeine intake indicated here may be slightly lower than summing caffeine intake from different sources (Table 13) because unknown values were set to zero here whereas in the previous table they were dropped from the calculation. The values were set to zero in this case to avoid dropping data due to missing values. It was discovered during follow-up calls that missing values tended to indicate occasional use only.

4.1.5 MTX intake by region

Tables 15 and 16 display the MTX intake from different sources and different constituents of MTX based on the geographic region in which the respondent lived. Because the sampling frame attempted to include a wide spectrum of regions in Southern Ontario, it is interesting to examine how much variation occurred between regions. For example, regular tea intake in Durham was twice that of York. However, coffee consumption was relatively similar among regions. The percentage of daily caffeine intake from regular coffee and regular tea varied as much as 12% among regions (see Table 15). The percentage of caffeine, theobromine and theophylline, total daily and lifetime intake, varied little by region (see Table 16).

of caffeine (in mg) and average percent of total caffeine intake by region.

	York		Halton-Peel		Ham-Went		Durham	
	mg	% caff	mg	% caff	mg	% caff	mg	% caff
Regular coffee	182	57.1	190[2]	55.4	178[1]	53.5	157	45.4
Instant coffee	27[1]	8.4	38[1]	11.1	24	7.2	42	12.1
Regular tea	51[2]	16.0	57[4]	16.6	73	21.9	98	28.3
Cola drinks	32[1]	10.0	34[3]	9.9	44	13.2	32[2]	9.2
Medications	4	1.3	12	3.5	4	1.2	10	2.9
Other (e.g.,choc bars)	23	7.2	12	3.5	10	3.0	7	2.0
Total	319	100.0	343	100.0	333	100.0	346	99.9

The numbers in superscript indicate the number of unknown values.

Table 16. Mean daily intake of different constituents of MTX (in mg) and average percent of total MTX intake by region.

	York		Halton-Peel		Ham-Went		Durham	
	mg	% MTX	mg	% MTX	mg	% MTX	mg	% MTX
Total caffeine[1]	318	93.0	339	92.3	332	91.0	344	89.6
Total theobromine	16	4.7	19	5.2	18	4.9	21	5.5
Total theophylline	8	2.3	9	2.5	15	4.1	19	4.9
Total methylxanthine	342	100.0	367	100.0	365	100.0	384	100.0

[1] The total caffeine intake indicated here may be slightly lower than summing caffeine intake from different sources (Table 15) because unknown values were set to zero here whereas in the previous table they were excluded from the calculation. The values were set to zero in this case was to avoid dropping data due to missing values. It was discovered during follow-up calls that missing values tended to indicate occasional use only.

4.1.6 MTX intake by smoking, alcohol and education level

There was no clear relationship between education level and coffee or caffeine intake across age and sex (see Table 17). Smokers (current and former) had a higher coffee and caffeine intake than individuals who had never smoked. The difference in intake between

coffee and caffeine intake than those who consumed alcohol.

Table 17. Daily mean coffee and caffeine intake (in mg) by education level, smoking history, and alcohol intake by age and sex.

	Male						Female					
	30-44 years		45-59 years		60-75 years		30-44 years		45-59 years		60-75 years	
	cof	caf	cof	caf	cof	caf	cof	caf	cof	caf	cof	caf
Education												
Grade 1-12	218	361	235	402	201	367	169	332	152	310	132	324
Uni/College	200	314	281	430	212	344	130	273	166	330	129	282
Smoker												
Never	174	289	196	325	208	346	100	239	127	269	114	263
Former	222	368	280	437	235	399	208	340	127	286	150	360
Current	255	369	266	446	172	224	131	318	298	523	132	333
Alcohol												
None	187	322	220	384	240	445	127	267	147	323	164	210
< 1 per day	214	328	230	403	128	308	135	289	123	287	162	366
1+ per day	193	319	329	456	277	376	201	320	302	450	120	250

4.2 Objective 2: Assess the misclassification of using coffee as a proxy measure for caffeine and MTX intake.

4.2.1 Kappa

Tables 18 and 19 show the exposure definitions used in the kappa analysis as well as the distribution of coffee and caffeine intake among these definitions. Repeat values made it difficult to set cutpoints for daily intake which would equally distribute values among categories. This difficulty was not encountered for lifetime intake.

Table 18. Exposure definitions for daily intake.

Caffeine (in mg)	Coffee cup equivalency	Distribution of coffee exposure (%)	Distribution of caffeine exposure (%)
0	0	32.9	1.5
1-114	0 - 1	4.2	7.5
115-229	1	19.3	6.1
230-344	2	20.1	19.5
345-459	3	13.0	26.4
460+	4 +	10.7	39.1

Table 19. Exposure definitions for lifetime intake.

Caffeine (in mg-yrs)	Distribution of coffee exposure (%)	Distribution of caffeine exposure (%)
0	21.8	0.2
0-2499	14.7	8.8
2500 - 4999	16.7	13.7
5000 - 7499	15.2	16.2
7500 - 12499	15.4	24.8
12500 +	16.2	36.3

Unweighted kappa values indicated no agreement between total daily coffee and total daily caffeine intake (k= -0.01), nor between total daily coffee and total daily MTX intake (k=0.00), and poor agreement for lifetime values (k=0.17, k=0.14, respectively). See Appendix G for worksheet.

Weighted kappa was used to account for the magnitude of discrepancy. Weighted kappa using squared error weights indicated poor levels of agreement (see Tables 20 and 21). The kappa values by age and sex of daily regular coffee and total daily caffeine intake varied from 0.38 to 0.50 and for lifetime intake from 0.30 to 0.61. Kappa values by age and sex for daily intake of regular coffee and total daily MTX intake were between 0.29 and 0.45. Agreement levels for lifetime intake of regular coffee and total daily MTX

were between 0.22 and 0.52. Results from the kappa statistic showed that coffee is an inadequate indicator for caffeine and MTX intake.

Table 20. Weighted kappa values (and standard errors) for daily caffeine intake from regular coffee and total daily caffeine intake by age and sex.

Stratum	Kappa present (s.e.)	Kappa lifetime (s.e.)
Males 30 - 44	.50 (.07)	.61 (.05)
Males 45 - 59	.41 (.07)	.33 (.03)
Males 60 - 75	.39 (.07)	.30 (.06)
Females 30 - 44	.43 (.10)	.54 (.11)
Females 45 - 59	.38 (.08)	.42 (.06)
Females 60 - 75	.40 (.08)	.32 (.04)
Total	.43 (.06)	.49 (.02)

Table 21. Weighted kappa values (and standard errors) for daily MTX intake from regular coffee and total daily MTX intake by age and sex.

Stratum	Kappa present (s.e.)	Kappa lifetime (s.e.)
Males 30 - 44	.45 (.08)	.52 (.04)
Males 45 - 59	.35 (.07)	.22 (.01)
Males 60 - 75	.33 (.06)	.26 (.06)
Females 30 - 44	.33 (.09)	.43 (.09)
Females 45 - 59	.29 (.06)	.33 (.04)
Females 60 - 75	.30 (.07)	.24 (.03)
Total	.34 (.03)	.41 (.01)

4.2.2 Misclassification matrices

Tables 22 and 23 display the misclassification matrices for the overall sample, that is the cross-tabulation of intake of caffeine from regular coffee by total caffeine intake. For the

daily and lifetime misclassification matrices by age and sex, refer to Appendix F. Many repeated values made it difficult to define exact quartiles for daily intake (see Table 22); exposure levels were more easily categorized by quartiles of lifetime intake of regular coffee (see Table 23).

The column totals show the distribution of total caffeine intake among exposure level categories. The row totals show the distribution of caffeine intake from regular coffee only. Individual cells along the left diagonal show the number of individuals whose exposure classification of total caffeine intake is in agreement with their exposure classification of caffeine intake from regular coffee only. The numbers above the left diagonal show the distribution of underestimation of caffeine from coffee among the different exposure levels. There were three missing values for cross-tabulation based on daily intake and 13 missing values for lifetime intake.

Of the 65 individuals whose total daily caffeine exposure was 1-115 mg of caffeine, 94% would have been classified as non-exposed based on their regular coffee intake. Of the 93 individuals whose total daily caffeine exposure was 116-230 mg of caffeine, based on their regular coffee intake, 2% would have been correctly classified, a further 48% would have been incorrectly classified in the lower adjacent exposure group (1-115 mg) and 49% would have been incorrectly classified as non-exposed, and so forth (see Table 22).

The misclassification for lifetime levels of exposure is somewhat lower than for daily intake but still pronounced (see Table 23). Misclassification percentages were very high and did not occur only among adjacent exposure levels.

Table 22. Cross-tabulation of daily caffeine intake from regular coffee (in mg) by total daily caffeine intake (in mg).

Level of exposure	Total caffeine				
Caffeine from coffee	0	1-115	116-230	231+	
0	7	61	46	43	157 (32.9%)
1-115	0	4	45	63	112 (23.4%)
116-230	0	0	2	94	96 (20.1%)
231+	0	0	0	113	113 (23.6%)
	7 (1.5%)	65 (13.6%)	93 (19.5%)	313 (65.5%)	478 (100.0%)

Table 23. Cross-tabulation of lifetime caffeine intake from regular coffee (in mg-yrs) by total daily caffeine intake (in mg-yrs).

Level of exposure	Total caffeine				
Caffeine from coffee	0	1-4150	4151-8999	9000+	
0	1	45	25	31	102 (21.8%)
1-4150	0	42	66	24	132 (28.2%)
4151-8999	0	0	41	74	115 (24.6%)
9000+	0	0	0	119	119 (25.4%)
	1 (0.2%)	87 (18.6%)	132 (28.2%)	248 (53.0%)	468 (100.0%)

4.3 Objective 3: To measure the effect of misclassification on odds ratio estimation.

While it is important to know how accurate a measure coffee is for total caffeine, it is also useful to measure how such misclassification would affect OR estimates, since misclassification may obscure an association.

4.3.1 Method 1: Hypothetical case-control distribution.

This method involved preselecting odds ratios estimates by constructing a hypothetical case-control distribution using true measures of caffeine intake and comparing the resulting OR estimates when coffee is used to approximate caffeine intake.

Setting the true ORs equal to 2.0 at all exposure levels for daily intake, using the non-exposed group as the referent, the resulting misclassified OR estimates were 1.0 for the overall sample at all levels of exposure, and the misclassified OR estimates ranged from 1.0 to 1.1 by age-sex stratum (see Table 24). When the hypothetical ORs were increased to 10.0, the misclassified OR estimates for the overall sample were 1.0 for all levels of exposure. For this scenario, the misclassified OR estimates by age and sex ranged from 1.0 to 1.1.

Applying a dose-response relationship to the hypothetical ORs, the resulting misclassified OR estimates had a much diluted dose-response relationship. At the two intermediate levels of exposure, the OR estimates were diluted to approximately 1.0, and in some cases were slightly below 1.0. At the highest level of exposure the misclassified OR was diluted from 3 to 1.2. All age-sex strata were affected similarly.

OR estimates for lifetime exposure are displayed in Table 26. In this case, different exposure categories were used to reflect coffee quartiles for each age-sex stratum (see Table 25). Misclassified lifetime OR estimates were similar to the daily misclassified OR estimates, but were slightly less biased.

The inclusion of all types of coffee in the definition, i.e., instant coffee, espresso, cappuccino and decaffeinated coffee, changed the results only slightly but, of course, introduced less bias than using regular coffee alone (see Tables 24, 26-28).

Table 24. True odds ratios (total daily caffeine intake) and misclassified odds ratios (daily caffeine intake from regular coffee) for hypothetical case-control study.

Exposure level	True OR	Misclassified OR						
		30-44 male	45-59 male	60-75 male	30-44 female	45-59 female	60-75 female	whole sample
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1-115	2.00	1.05	1.00	1.03	1.03	1.00	1.03	1.02
116-230	2.00	1.05	1.00	1.03	1.03	1.00	1.03	1.02
> 230	2.00	1.05	1.00	1.03	1.03	1.00	1.03	1.02
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1-115	10.00	1.09	1.00	1.06	1.06	1.00	1.05	1.04
116-230	10.00	1.09	1.00	1.06	1.06	1.00	1.05	1.04
> 230	10.00	1.09	1.00	1.06	1.06	1.00	1.05	1.04
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1-115	1.50	0.96	0.88	0.91	0.93	0.88	0.88	0.90
116-230	2.00	1.01	0.91	0.96	0.96	1.02	1.07	0.99
> 230	3.00	1.20	1.10	1.15	1.25	1.18	1.24	1.19
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1-115	0.80	1.12	1.46	1.28	1.17	1.34	1.29	1.29
116-230	0.50	0.89	1.22	1.02	0.98	0.90	0.78	0.94
> 230	0.20	0.52	0.69	0.58	0.48	0.55	0.48	0.54

Table 25. Exposure level definitions for lifetime of caffeine intake estimates (in mg-yrs) for each stratum.

	Exposure level			
	1	2	3	4
Males 30-44	0-999	1000-3199	3200-5999	6000+
Males 45-59	0-199	200-6699	6700-11699	11700+
Males 60-75	0-5999	6000-10999	11000-16999	17000+
Females 30-44	0	1-1799	1800-3999	4000+
Females 45-59	0-999	1000-3999	4000-7999	8000+
Females 60-75	0	1-5799	5800-12399	12400+
Whole sample	0	1-4499	4500-9499	9500+

Table 26. True odds ratio (total lifetime caffeine intake) and misclassified odds ratio (lifetime caffeine intake from regular coffee) in a hypothetical case-control study.

Exposure level*	True OR	Misclassified OR						
		30-44 male	45-59 male	60-75 male	30-44 female	45-59 female	60-75 female	whole sample
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	2.00	1.11	1.03	1.11	1.00	1.02	1.02	1.00
3	2.00	1.11	1.03	1.11	1.00	1.02	1.02	1.00
4	2.00	1.11	1.03	1.11	1.00	1.02	1.02	1.00
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	10.00	1.22	1.06	1.22	1.00	1.04	1.04	1.01
3	10.00	1.22	1.06	1.22	1.00	1.04	1.04	1.01
4	10.00	1.22	1.06	1.22	1.00	1.04	1.04	1.01
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	1.50	0.93	0.85	0.97	0.80	0.95	0.93	0.85
3	2.00	1.10	0.88	1.14	0.91	0.96	0.95	0.99
4	3.00	1.40	1.27	1.33	1.23	1.20	1.20	1.24
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0.80	1.12	1.13	1.06	1.54	1.15	1.20	1.39
3	0.50	0.89	1.14	0.68	1.11	1.02	1.03	1.02
4	0.20	0.52	0.48	0.42	0.51	0.54	0.53	0.50

* See Table 25 for exposure level definitions for each age-sex stratum.

Table 27. True odds ratio (total daily caffeine intake) and misclassified odds ratio (daily caffeine intake from all types of coffees combined*) for hypothetical case-control study.

Exposure level	True OR	Misclassified OR						
		30-44 male	45-59 male	60-75 male	30-44 female	45-59 female	60-75 female	whole sample
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1-115	2.00	1.07	1.05	1.08	1.05	1.00	1.08	1.05
116-230	2.00	1.07	1.05	1.08	1.05	1.00	1.08	1.05
> 230	2.00	1.07	1.05	1.08	1.05	1.00	1.08	1.05
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1-115	10.00	1.14	1.10	1.15	1.09	1.00	1.15	1.10
116-230	10.00	1.14	1.10	1.15	1.09	1.00	1.15	1.10
> 230	10.00	1.14	1.10	1.15	1.09	1.00	1.15	1.10
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1-115	1.50	0.81	0.86	0.93	0.81	0.85	0.88	0.85
116-230	2.00	1.05	0.99	1.01	1.01	0.92	1.09	1.01
> 230	3.00	1.23	1.16	1.22	1.28	1.18	1.31	1.23
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1-115	0.80	1.47	1.42	1.19	1.43	1.44	1.24	1.38
116-230	0.50	0.82	0.94	0.89	0.88	1.09	0.75	0.89
> 230	0.20	0.49	0.56	0.50	0.46	0.55	0.43	0.49

* Regular, instant, espresso, cappuccino and decaffeinated coffee.

Table 28. True odds ratio (total lifetime caffeine intake) and misclassified odds ratio (lifetime caffeine intake from all types of coffees combined*) for hypothetical case-control study.

Exposure level**	True OR	Misclassified OR						
		30-44 male	45-59 male	60-75 male	30-44 female	45-59 female	60-75 female	whole sample
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	2.00	1.14	1.15	1.15	1.05	1.15	1.17	1.18
3	2.00	1.14	1.15	1.15	1.05	1.15	1.17	1.18
4	2.00	1.14	1.15	1.15	1.05	1.15	1.17	1.18
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	10.00	1.27	1.31	1.32	1.10	1.31	1.36	1.38
3	10.00	1.27	1.31	1.32	1.10	1.31	1.36	1.38
4	10.00	1.27	1.31	1.32	1.10	1.31	1.36	1.38
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	1.50	0.82	1.08	1.01	0.87	1.04	1.07	0.98
3	2.00	1.16	0.89	1.16	1.00	1.09	1.09	1.11
4	3.00	1.50	1.55	1.42	1.32	1.43	1.47	1.55
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0.80	1.27	0.89	0.99	1.27	0.94	0.90	1.04
3	0.50	0.69	1.00	0.68	0.90	0.78	0.77	0.75
4	0.20	0.34	0.34	0.37	0.43	0.37	0.36	0.33

* Regular, instant, espresso, cappuccino and decaffeinated coffee.

** See Table 25 for exposure level definitions for each age-sex stratum.

Method 2: Odds ratio corrections to previous studies.

The second method of studying the effects of misclassification on risk involved "correcting" OR estimates reported in the literature. Two articles were chosen which used coffee as the variable of interest: one involved daily coffee intake estimates before illness (MacMahon *et al.,* 1981) and the other involved lifetime estimates (Jain *et al.*, 1991). Although the crude results from these studies are adjusted here, this does not presume that they used coffee as a surrogate measure for caffeine intake. Each article did indeed report findings based on coffee intake only. These corrections are for the purpose of assessing the effects of equating coffee with caffeine.

Estimates of the true population of MacMahon *et al.*'s (1981) study on coffee yielded values that were inadmissible, partly because the sensitivity was low but also because the distribution of the data for these cutpoints was unequal (see Tables 29 and 30). More evenly spread data between categories would have resulted in positive numbers. Results from the correction algorithm applied to Jain *et al.* (1991) also yielded inadmissible values for multiple levels of exposure (see Tables 32 and 33). When collapsing the three lowest levels of exposure to form a 2x2 table, both the crude and corrected odds ratio estimates were equal to 1.0 (see Tables 34 and 35). This study had a lower consumption of coffee than either of these two studies (see Tables 31 and 36).

Table 29. Distribution of cases and control by coffee intake in MacMahon *et al.* (1981) and crude odds ratio estimates.

	Coffee intake (cups/day)				Total
	0	1-2	3-4	5+	
Cases (n)	20	153	106	88	367
Controls (n)	32	119	74	82	307
Crude OR	1.0	2.1	2.3	1.7	

Table 30. Distribution of cases and control by coffee intake in MacMahon *et al.* (1981) applying correction algorithm.

	Coffee intake (cups/day)				Total
	0	1-2	3-4	5+	
Cases (n)	-136	-36	107	432	367
Controls (n)	-91	-35	31	403	307

Table 31. Percent distribution of the coffee intake in the MacMahon *et al.* (1981) study population and the present study.

	Coffee intake (cups/day)				Total
	0	1-2	3-4	5+	
Present study	37.4	39.1	18.9	4.6	100.0
MacMahon *et al.* controls	10.4	38.8	24.1	26.7	100.0

Table 32. Distribution of cases and control by coffee intake in Jain *et al.* (1991) and crude odds ratio estimates.

	Coffee intake (cup-years)				Total
	None	0-39	40-110	110+	
Cases (n)	25	69	76	76	246
Controls (n)	40	136	174	154	504
Crude OR	1.0	1.3	1.0	1.2	

Table 33. Distribution of cases and control by coffee intake in Jain *et al.* (1991) applying correction algorithm.

	Coffee intake (cup-years)				Total
	None	0-39	40-110	110+	
Cases (n)	-41	47	86	155	247
Controls (n)	-100	84	205	314	503

Table 34. Crude odds ratio estimates in a collapsed 2x2 table.

	Coffee intake (cup-years)		Total
	0-110	110+	
Cases (n)	170	76	246
Controls (n)	350	154	504
Crude OR	1.0	1.0	

Table 35. Corrected odds ratio estimates in a collapsed 2x2 table.

	Coffee intake (cup-years)		Total
	0-110	110+	
Cases (n)	92	155	247
Controls (n)	189	314	503
Corrected OR	1.0	1.0	

Table 36. Percent distribution of the coffee intake in the Jain *et al.* (1991) study population and the present study.

	Coffee intake (cup-years)				Total
	None	0-39	40-110	110+	
Present study	23.9	24.5	31.4	20.2	100.0
Jain *et al.* controls	7.9	27.0	34.5	30.6	100.0

5. DISCUSSION

First, this section will summarize and interpret the most important findings from the three study objectives: 1) to describe patterns of coffee, caffeine and MTX intake, 2) to determine the misclassification caused by using coffee as a proxy measure for caffeine and MTX, and 3) to measure the effects of this misclassification on the odds ratio estimates. The third objective was the main goal of the study; it assessed the extent to which the misclassification would dilute a true relationship between caffeine exposure and disease. The remainder of this section will describe the significance of this study, its limitations and offer recommendations for future studies.

5.1 Interpretation of results

5.1.1 Objective 1: To estimate coffee, caffeine and MTX intake.

The percentage of men in the Ontario Health Survey (OHS) aged 30-75 who reported coffee and tea use did not differ from this study sample by more than 5% (see Table 7). The differences in reporting between the two studies were more pronounced for women. In the OHS, three quarters of women aged 30-75 drank coffee, whereas approximately two thirds did so in this sample. The situation was reversed concerning tea consumption:

thirds of the women aged 30-75 in the OHS reported regular use. The average mean consumption of tea and coffee was higher in this sample than in the OHS. In contrast, coffee consumption was lower in this sample than what was reported by the Coffee Association. Further, the levels of coffee exposure were much lower in this study sample than in either of the two corrected articles (see Tables 31 and 36).

This study found that overall, 54% of daily caffeine intake was from regular coffee; 20% from tea; 11% from soft drinks; 10% from instant coffee and 5% from all other sources combined (refer to Table 10, Figure 5 for breakdown by age-sex stratum). Based on data reported by Statistics Canada in the 1980's, it was expected that 55% of caffeine would be obtained from coffee, 32% from tea, 7% from soft drinks and 1% from chocolate (Gilbert, 1984).

There were no Canadian data with which to compare this sample's caffeine intake by age and sex, therefore American data were used as an approximation. Contrary to what was estimated in the United States (Pozniak, 1985), females in this study reported lower caffeine intake than males (see Table 14). However, similar to the United States, it was found that individuals aged 45-59 had the highest caffeine intake of any other age group, both in males and in females.

Briefly, the five key findings for daily intakes were:

1. Coffee was the main source of caffeine for all ages, both for males and females. Tea was the second highest source for females of all ages and for males aged 60-75. Tea was a close second to coffee for females aged 60-75. Cola soft drinks were the second highest caffeine source for males aged 30-44 and, for males aged 45-59 the level of consumption of caffeine from soft drinks was equal to that of tea. Although the reported number of servings of tea per day surpassed the reported number of cups of coffee per day for females aged 45-59 and females aged 60-75, tea contains slightly less than half the amount of caffeine of coffee, and therefore tea never represented a greater source of caffeine than did coffee (Table 11, Figure 5).
2. The amount of regular coffee consumed varied by age and sex: males reported a higher intake than did females; both males and females aged 45-59 reported the highest coffee intake. The amount of regular tea also varied by age and sex. In this case, females reported a higher consumption than males; both males and females aged 60-75 reported the highest tea intake (see Table 11, Figure 5). Instant coffee and cola soft drink intake also varied by age and sex.
3. In descending order, regular coffee, regular tea, cola soft drinks and instant coffee contributed on average 94.1% of caffeine intake. The contribution of other caffeine

beverages (e.g., chocolate milk, hot cocoa), medications (e.g. Midol, 222s, Cafergot) and chocolate bars was less than 2% from each source (see Table 11, Figure 5).

4. The vast majority (91.6%) of MTX was in the form of caffeine; 5.1% was from theobromine and 3.3% was from theophylline (see Table 12). It should be noted that conservative MTX values for both theobromine and theophylline were used.
5. There was some variation in caffeine intake by region, for example, that residents of Durham reported a higher tea intake than did those of other regions and this was not accounted for by age and sex differences (see Tables 15 and 16).

It should be noted that lifetime estimates followed similar patterns to daily estimates for the findings discussed in the above five points (see Tables 13 and 14, Figure 6).

5.1.2 Objective 2: To determine the misclassification of using coffee as a proxy measure for caffeine and MTX

Indices of the accuracy of measurement

One way to examine the accuracy of using coffee as an index for caffeine is to calculate the percentage of caffeine provided by coffee. Coffee provided between 42 - 63% of total

caffeine intake, with females aged 60-75 at the lowest end of the range and males aged 30-44 at the highest. From these findings, it appears that using average coffee intake is not a complete estimate for average caffeine intake since it provides only on average 53.3% of caffeine intake.

Misclassification matrix

A second and more detailed way of measuring the accuracy of using regular coffee as an index for caffeine is using misclassification matrices. Here, these matrices were constructed by cross-tabulation of true caffeine intake by the classification of caffeine that would result from regular coffee only (see Methods for full explanation, Figure 4). For daily intake, both coffee and caffeine exposure levels were defined based on coffee intake quartiles by caffeine content (mg) in: 0 cups, 0-1 cups, 2 cups, or 3 cups or more of coffee. These matrices can be found in Tables 22 and 23 and in Appendix F.

The underestimation of caffeine intake by using caffeine from only regular coffee was much higher than expected. The exposure level of caffeine intake from all sources was classified in the correct category based on their daily caffeine intake from regular coffee

4% of those reporting no current regular coffee intake had no current caffeine intake. Of the remaining 96% of non-coffee drinkers, 39% were in the second quartile of exposure; 29% in the third; and 27% in the highest quartile. These findings indicate that the vast majority of individuals would be underestimated in their caffeine intake *and* that the underestimation is not only in adjacent exposure level categories as it is assumed in misclassification literature (Marshall *et al.*, 1990). Similar patterns of underestimation occurred for lifetime intake.

The distribution of underestimation was not identical among the age and sex strata. Although males aged 30-44 reported the highest percentage of caffeine from coffee (62.8%), their underestimation did not appear to be less than the other age-sex strata. Similarly, females aged 60-75, with the lowest percentage caffeine intake from regular coffee, showed no greater underestimation than other age-sex groups. Therefore, assessing how good an index regular coffee is for caffeine intake is a more complex task than simply calculating the percentage of caffeine intake obtained from coffee.

Here, Kappa statistics provide a summary value of the strength of agreement between the classification of regular coffee and total caffeine. The calculated values confirmed what the misclassification matrix strongly suggested: coffee is a poor surrogate measure for caffeine intake. Kappa values for both daily and lifetime estimates were in the poor or slightly above poor agreement range (see Table 20). This was true for all age-sex strata, except for males and females aged 30-45 whose kappa values were in the midrange. Because of the high number of categories used in the weighted analyses, the kappa values will be higher than had fewer categories been used (Brenner & Kliebsch, 1996).

Kappa values were also calculated to test the agreement between coffee intake and total MTX intake (see Table 21). As expected, Kappa values were slightly lower than those of caffeine. Any bias found from using coffee for caffeine estimates would be higher for MTX intake because MTX encompasses two other substances, theobromine and theophylline, in addition to caffeine.

MTX was excluded as an outcome of interest in further analyses due to measurement issues. Basically, theobromine content is highly variable, particularly in chocolate products. Estimates of theophylline were adequately assessed by measuring regular tea,

theobromine would have been required to make accurate MTX estimates.

5.1.3 Objective 3: To measure the effects of underestimation on the odds ratio

The main purpose of measuring caffeine intake and constructing classification matrices was to determine the effects of such underestimation on odds ratio estimates for case-control studies. This was achieved using two methods: 1) constructing a hypothetical case-control distribution set to pre-determined odds ratios and 2) correcting actual data from previous studies.

Method 1: hypothetical case control distribution

A hypothetically increased risk of caffeine intake of 10.0 with four levels of exposure resulted in a substantial dilution of the degree of association. The odds ratio estimates, using caffeine from regular coffee only, were reduced from 10.0 to no more than 1.1 at any exposure level for all age-sex strata. Typically, epidemiological studies do not expect to find such extreme associations. This would mean that assuming that caffeine is mainly obtained from coffee may completely obscure a potentially increased risk.

differences in factors that affect the amount of bias, such as the misclassification rates and the distribution of subjects among exposure levels (Birkett, 1992). The effects of misclassification were the highest for males and females aged 45-59, in which the association was completely diluted (OR=1.0). Results were similar between daily and lifetime intakes, although lifetime misclassification effects were slightly less pronounced (see Tables 24 and 26). Therefore, if caffeine is the variable of interest and coffee is used as an index for caffeine use, it may not be surprising that some studies have found no effect.

If the true nature of the relationship between caffeine and a certain disease was a dose-response relationship rather than a threshold, the results would likely also be obscured. True ORs set at 1.5, 2.0, 3.0 for each exposure level generally produced misclassified OR estimates slightly below 1.0 at two intermediate levels of exposure and an OR estimate of no more than 1.3 at the highest level of exposure. This was the case for daily and lifetime intake.

In order to have an accurate measure of caffeine intake, it would probably be sufficient to measure only the four main sources of caffeine, that is, regular coffee, regular tea, cola soft drinks and instant coffee. In this sample, these captured 90.1% to 98.3% of caffeine

intake and resulted in misclassification estimates that did not affect the odds ratio more than 0.1 in any age sex-strata (data not shown).

Method 2: odds ratio corrections to previous studies

It was not possible to adjust odds ratio estimates from published case-control studies on coffee and pancreatic cancer in multiple levels of exposure to reflect estimates based on total caffeine intake; the underestimation was too extreme. For the Jain *et al.* (1991) study, it was possible to obtain values for a 2x2 table by collapsing the three lowest multiple exposure levels; the odds ratio was 1.0 for both crude and corrected estimates.

Comparison of method 1 and method 2

The advantage of using the hypothetical case-control construction is that it makes fewer assumptions than applying corrections to previous studies. The hypothetical case-control distribution uses the misclassification matrices to develop correction algorithms. These are then applied to true ORs to obtain misclassified OR estimates. The misclassification matrix is constructed from consumption patterns from this sample. Using these matrices in order to correct previous studies assumes that 1) exposure prevalence from this study is similar to the one study found in the literature and 2) misclassification rates of using

to make as this sample reported much lower intake than either of the two cited studies (see Tables 34 and 36).

5.2 Limitations

Validity of recall

Accurate recall is a potential problem in all dietary studies. Although reported coffee consumption has been found to be less variable from day to day than reported food consumption (Wu *et al.*, 1988), it is perhaps subject to some problems of recall. Lifetime recall may be less accurate than current patterns. Whenever assessing past dietary patterns, memory is strongly influenced by current patterns (Wu *et al.*, 1988) and the greater length of time the less accurate the recall (Biemer *et al.*, 1991). Insofar as recall may be an issue, recall inaccuracies would have to alter the relationship of coffee to caffeine sufficiently to influence the corrections to the odds ratio in order to affect the results of this study; there is no reason to suggest that this occurred.

The questionnaire itself was limited by a key factor inherent in MTX studies: MTX variability. It is exceedingly difficult to obtain accurate MTX measurement due to the many factors affecting MTX content, such as brand used, strength of brew, brewing method and volume in MTX-containing beverages.

There is often a trade-off made between gathering the most detailed information possible and making the questionnaire fast and easy to complete. In this case, information could have been collected on changing consumption over time. The questionnaire simply asked what age the person started drinking coffee, the usual number of servings, and the age at which he or she stopped (if applicable). Thus, the assumption was that either the person had a consistent intake or entered an intake value averaged over time. It would have been preferable to ask if the consumption had increased, decreased or remained stable over time.

Another weakness of the questionnaire was that it did not collect information on serving sizes that were between 0 and 1. The way the question was formulated allowed only a number of servings for daily use; individuals who consumed a beverage on a less regular

basis had no clear way of answering the question. It would have been better to ask individuals to specify the time period of consumption, e.g. daily, weekly or monthly and to ask if the standard volume was a cup or a mug (see Schreiber, 1988), considering that serving size varies among individuals (Stavric *et al.*, 1988). Finally, insufficient information was gathered on chocolate products -- the main source of theobromine -- to provide accurate theobromine measures.

An issue that might affect caffeine consumption levels in the future is decaffeinated cola soft drinks. The beverage industry reported that, in 1994, approximately 5% of all cola soft drinks sold were in the decaffeinated format (Barone & Roberts, 1996). Also, while espresso and cappuccino were included in the list of beverages, café au lait and cafe latte, which are becoming increasingly popular, were not included. Distinctions also need to be made between caffeinated and decaffeinated specialty coffees such as these.

On a final note, the following quote from Spiller (1984, p.3) cautions that combining all sources of MTXs may be as subject to problems as studying them separately.

> We must always remember that these products supply a *host* of substances other than MTX. We cannot necessarily equate the effects of drinking 70 mg of caffeine in a cup of coffee with the same amount in either tea or a pharmaceutical preparation, the caffeine may be modified in each case by accompanying compounds.

This study clearly showed that coffee and caffeine are not equivalent exposures. Results suggest that using coffee as a surrogate measure for caffeine masks any true association regardless of age, sex, type of relationship (threshold or dose-response) for daily and lifetime exposure. Studies examining the relationship between coffee and disease status have found only weak associations which have been inconsistent from study to study. The findings from this study demonstrated that measuring coffee instead of caffeine may contribute to the lack of positive findings.

For future studies measuring total caffeine intake, measuring the four main sources of intake which were regular coffee, tea, soft drinks and instant coffee would probably be sufficient. A more detailed assessment of theobromine and theophylline would be appropriate in order to secure a better measurement of MTX.

This study showed the considerable effects of misclassification that were apparent from examining only one aspect of a host of potential measurement problems. If one studied the factors relevant to many dietary studies that influence absorption, such as tobacco and alcohol use, diet, individual's weight and height, then the measurement problems would potentially be greater. Considering the further difficulty in obtaining accurate exposure

and chocolate, in addition to the wide daily fluctuation due to different preparation methods, one is left wondering how reasonable it is to expect to detect a potentially increased risk.

Armstrong BK, White E, Saracci R (1992). Principles of Exposure Measurement in Epidemiology. Oxford University Press, New York, 351 pp.

Association Scientifique Internationale du café (1993). The 15th International Scientific Colloquium on Coffee, Montpellier, 6-11 June, 1993, volume II, 902 pp.

Barone JJ, Grice HC (1990). Sixth International Caffeine Workshop, Hong Kong, 7-10 August 1989. *Food and Chemical Toxicology*, 28: 279-283.

Barone JJ, Roberts HR (1996). Caffeine Consumption. *Food and Chemical Toxicology, 34(1): 119-129.*

Biemer PP, Groves RM, Mathiowietz NA, Sudman S (eds) (1991). Measurement Errors in Surveys. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.

Birkett NJ (1992). Effect of Nondifferential Misclassification on Estimates of Odds Ratios with Multiple Levels of Exposure. *American Journal of Epidemiology*, 136: 356-362.

Bland JM, Altman DG (1986). Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *The Lancet*: 307-310.

Brenner H (1991). Does Nondifferential Misclassification of Exposure Always Bias: A True Effect Toward the Null Value? *American Journal of Epidemiology*, 134(4): 438-440.

Brenner H, Kliebsch U (1996). Dependence of Weighted Kappa Coefficients on the Number of Categories. *Epidemiology*, 7(2): 199-202.

Bueno de Mesquita HB, Smeets FWM, Runia S, Hulshof KFAM (1992). The Reproducibility of a Food Frequency Questionnaire Among Controls Participating in a Case-Control Study on Cancer. *Nutrition and Cancer*, 18(2): 144-156.

Bullough BB, Hindi-Alexander M, Fetouh S (1990). Methylxanthines and Fibrocystic Breast Disease: A Study of Correlations. *Nurse Practitioner*, 15(3): 36-44.

Canadian Pharmaceutical Association (1995) Compendium of Nonprescription Products.

Canadian Pharmaceutical Association (1995) Compendium of Pharmaceuticals and Specialities. 35th edition

Chou T (1992). Wake Up and Smell the Coffee: Caffeine, Coffee, and the Medical Consequences. *The Western Journal of Medicine*, 157: 544-553.

Clavel F, Benhamou E, Auquier A, Tarayre M, Flamant R (1989). Coffee, Alcohol, Smoking and Cancer of the Pancreas: A Case-Control Study. *International Journal of Cancer*, 43: 17-21.

Coffey RJ, Go VLW, Zinsmeister AR, DiMagno EP (1986). The Acute Effects of Coffee and Caffeine on Human Interdigestive Exocrine Pancreatic Secretion. *Pancreas*, 1(1): 55-61.

Copeland KT, Checkoway H, McMichael AJ, Holbrook RH (1977). Bias Due to Misclassification in the Estimation of Relative Risk. *American Journal of Epidemiology*, 105(5): 488-495.

Curatolo PW, Robertson D (1983). The Health Consequences of Caffeine. *Annals of Internal Medicine*, 98 (part I): 641-653.

De Planer A (1989). Bowes & Church's Food values. Lipincott: 15th edition.

D'Amicis A, Viani R (1993). The Consumption of Coffee. In *Caffeine, Coffee and Health* Edited by S Garattini. New York: Raven Press, pp 1-16.

Debry G (1993). Le café et la santé. Paris: John Libey, 188 pp.

Dosemeci M, Wacholder S, Lubin JH (1990). Does Misclassification Always Bias True Effect Toward The Null Value? *American Journal of Epidemiology*, 132: 746-748.

Ensminger A, Ensminger ME, Karerde JE, Robson JR (eds) (1994). Food and Nutrition, 2nd Edition.

Falk RT, Pickle LW, Fontham ET, Correa P, Fraumeni JF (199088?). Life-Style Risk Factors for Pancreatic Cancer in Louisiana: A Case-Control Study. *American Journal of Epidemiology*, 128(2): 324-336.

and the Use of Tobacco, Alcohol and Coffee. *International Journal of Cancer*, 45: 816-820.

Fleiss JL (1981). The Measurement of Interrater Agreement. *In Statistical Methods for Rates and Proportions*, 2nd Edition. New York: Wiley and Sons, pp. 212-239.

Fontham ETH, Correa P (1989). Epidemiology of Pancreatic Cancer. *Surgical Clinics of North America*, 69(3): 551-567.

Fredholm (1984). Gastrointestinal and Metabolic Effects of Methylxanthines. In *Methylxanthine Beverages and Foods: Chemistry, Consumption, and Health Effects.* Edited by GA Spiller. New York: Alan R. Liss, inc., pp. 303-354.

Garattini S (1993). Caffeine, Coffee and Health. Monographs of the Mario Negri Institute for Pharmacological Research. Raven Press: New York, 420 pp.

Ghadirian P, Simard A, Baillaregeon J (1991). Tobacco, Alcohol, and Coffee and Cancer of the Pancreas. *Cancer*, 67, 2664-2670.

Gilbert RM (1985). Caffeine Consumption. In Methylxanthine Beverages and Foods: Chemistry, Consumption, and Health Effects. Edited by GA Spiller. New York: Alan R. Liss, inc., pp. 185-213.

Glass RM (1994). Caffeine Dependence: What Are the Implications? *JAMA* 272(13): 1065-1066.

Gordis L (1990). Consumption of Methylxantine-Containing Beverages and Risk of Pancreatic Cancer. *Cancer Letters*, 52: 1-12.

Graham HN (1984). Tea: The Plant and Its Manufacture. Chemistry and Consumption of the Beverage. In Methylxanthine Beverages and Foods: Chemistry, Consumption, and Health Effects. Edited by GA Spiller. New York: Alan R. Liss, inc., pp. 29-74.

Graham P, Jackson R (1993). The Analysis of Ordinal Agreement Data: Beyond Weighted Kappa. *Journal of Clinical Epidemiology*, 46(9): 1055-1062.

Guggenmoos-Holzmann (1993). How Reliable are Chance-Corrected Measures of Agreement. *Statistics in Medicine*, 12: 2191-2205.

in chapter 11 of Spiller GA ed. Methylxanthine Beverages and Foods: Chemistry, Consumption, and Health Effects. *Progress in Clinical and Biological Research*, volume 158. Alan R. Liss, inc., New York, 413 p.

Howe GR, Jain M, Miller AB (1990). Dietary Factors and Risk of Pancreatic Cancer: Results of a Canadian Population-Based Case-Control Study. *International Journal of Cancer*, 45: 604-608.

Hsieh CC, MacMahon B, Yen S, Trichopulous D, Warren K, Nardi G (1986). Coffee and Pancreatic Cancer. *The New England Journal of Medicine*, 315: 587-589.

Hui, YH (1994). Encyclopedia of Food Science and Technology. John Wiley & Sons, Inc.: New York.

International Agency for Research on Cancer (1991). IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 51, Coffee, Tea, Mate, Methylxanthines and Methylglyoxal, 513 pp.

Jain M, Howe GR, St. Louis P, Miller AB (1991). Coffee and Alochol as Determinants of Risk of Pancreas Cancer: A Case-Control Study from Toronto. *International Journal of Cancer*, 47: 384-389.

James JE (1991). Caffeine and Health. Academic Press: London, 432 pp.

Kleinbaum DG, Kupper LL, Morgenstern H (1982). Information Bias. In Epidemiologic Research. New York: Van Nostrand-Linehold, 220-239.

Kramer HC, Bloch DA (1988). Kappa Coefficients in Epidemiology: An Appraisal of a Reappraisal. *Journal of Clinical Epidemiology*, 41(10): 959-968.

Lamarine RJ (1994). Selected Health and Behavioural Effects Related to the Use of Caffeine. *Journal of Community Health*, 19(6): 449-466.

Lander N, Soloway AH, Minton JP, Rawal BD, Gairola C (1988). Potential Metabolic Mutagens of Caffeine and Various Methylxanthines. *Journal of Pharmaceutical Sciences*, 77(11): 955-958.

La Vecchia C (1990). Epidemiological Evidence on Coffee and Digestive Tract Cancers: A Review. *Dig Dis*, 8: 281-286.

La Vecchia C, Negri E, Franceschi S, D'Avanzo B, Boyle P (1992). Tea Consumption and Cancer Risk. *Nutrition and Cancer*, 17(1): 27-31.

Lima DR (1990). Coffee a Medicinal Plant: All about Coffee and Its Effect on Human Health. Vantage Press: New York, 155 pp.

Lubin F, Ron E (1990). Consumption of Methylxanthine-Containing Beverages and the Risk of Breast Cancer. *Cancer,* 53: 81-90.

Lyon JL, Egger MJ, Robison LM, French TK, Gao R (1992). Misclassification of Exposure in a Case-Control Study: The Effects of Different Types of Exposure and Different Proxy Respondents in a Study of Pancreatic Cancer. *Epidemiology*, 3(3): 223-231.

Lyon JHL, Mahoney AW, French TK, Moser R (1992). Coffee Consumption and the Risk of Cancer of the Exocrine Pancreas: A Case-Control Study in a Low-Risk Population. *Epidemiology*, 3(2): 164-170.

Mack, TM, Yu MC, Hanisch R, Henderson BE (1986). Pancreas Cancer and Smoking, Beverage Consumption, and Past Medical History. *Journal of the National Cancer Institute*, 76: 49-60.

Maclure M, Willet WC (1987). Misinterpretaion and Misuse of Kappa Statistic. *American Journal of Epidemiology*, 126: 161-169.

MacMahon B, Yen S, Trichopoulos D, Warren K, Nardi G (1981). Coffee and Cancer of the Pancreas. *The New England Journal of Medicine*, 304(11): 630-633.

Marshall RJ (1994) Misclassification of Exposure in Case-control studies: Assessment by Quality Indices. *Epidemiology*, 5: 309-314.

Marshall RJ (1990). Validation Study Methods for Estimating Exposure Proportions and Odds Ratios with Misclassified Data. *Journal of Clinical Epidemiology*, 43(9): 941-946.

McLaughlin, CC, Mahoney MC, Nasca PC et al. (1992). Breast Cancer and Methylxanthine Consumption. *Cancer Causes and Control,* 3: 175-178.

Ministry of Health, Ontario Health Survey (1990). User's Guide, volume Microdata Manual. Analysis provide by M. Agha at ICES, Toronto.

Myers MG, Basiniski A (1992). Coffee and Coronary Heart Disease. *Arche Intern Med*, 152: 1767-1772.

National Soft Drink Association. Internet: http://www.seas.upen.ed/~cpage/caffeine/FAQMain.html. Site visited on November 12, 1996.

Nehlig A, Derby G (1994). Effets du café et de la caféine sur la fertilité, la lactation et le développement. *Journal Gynecol. Obstet. Biol. Repr.*, 23: 241-256.

Olsen GW, Mandel JS, Gibson RW, Wattenberg LW, Schuman LM (1989). A Case-Control Study of Pancreatic Cancer and Cigarettes, Alcohol, Coffee and Diet. *American Journal of Public Health*, 79(8): 1016-1019.

Pafai and Jankiewicz (1991). Physicans Desk Reference and Distribution of Food Technologies. Drug and Human Behaviour.

Peacock JL, Bland JM, Anderson HR (1991). Effects on Birthweight of Alcohol and Caffeine Consumption in Smoking Women. *Journal of Epidemiology and Community Health*, 45: 159-163.

Pozniak PC (1985). The Carcinogenicity of Caffeine and Coffee: A review. *Journal of the American Dietetic Association*, 85(9): 1127-1133.

Puccio EM, McPhillips JB, Barrett-Connor E, GaniatisTG (1990). Clustering of Atherogenic Behaviors in Coffee Drinkers. *American Journal of Public Health,* 80 (11): 1310-1313.

Roberts I, Lee-Joe T (1993). Effect of Exposure Measurement Error in a Case-Control Study of Child Pedestrian Injuries. *Epidemiology*, 4(5): 477-479.

Rohan TE, Cook MG, McMichael AJ (1989). Methylxanthines and Benign Proliferative Epithelial Disorders of the Breast in Women. *International Journal of Epidemiology*, 18(3): 626-633.

with Error. *Statistics in Medicine*, 15: 293-303.

SAS® Institute Inc., Box 8000, Cary. NC 27512-8000, USA.

Schreiber GB, Robins M (1988). Confounders Contributing to the Reported Association of Coffee or Caffeine With Disease. *Preventive Medicine*, 17: 280-309.

Sosenko JM, Gardner LB (1987). Attribute Frequency and Misclassification Bias. *Journal of Chronic Disiseases*, 40(3): 203-207.

Spiller GA ed (1985). The Methylxanthine Beverages and Foods: Chemistry, Consumption, and Health Effects. *Progress in Clinical and Biological Research*, volume 158. Alan R. Liss, inc., New York, 413 p.

Spiller MA ed (1985). The Coffee Plant and Its Processing. In The Methylxanthine Beverages and Foods: Chemistry, Consumption, and Health Effect. Edited by GA Spiller, 75-90 pp.

Statistics Canada, (1996). CANSIM, Matrix 63967. Internet site: http://www.statscan.ca/english/Pgdb/People/population. Site visited on May 7, 1996.

Stavric B, Klassen R, Watkinson B, Karpinski K, Stapleu R (1988). Variability in Caffeine Consumption from Coffee and Tea: Possible Significance for Epidemiological Studies. *Food Chemistry and Toxicology*, 26(2): 111-118.

Stavric B (1992). An Update on Research with Coffee/Caffeine. *Food and Chemical Toxicology*, 30(6): 535-555.

Thompson WD, Walter SD (1988). Variance and Dissent: A Reappraisal of the Kappa Coefficient. *Journal of Clinical Epidemiology*, 41(10): 949-958.

Thompson WG (1994) Coffee: Brew or Bane? *The American Journal of the Medical Sciences*, 308(1): 49-57.

Viscoli CM, Lachs MS, Horowitz RI (1993). Bladder Cancer and Coffee Drinking: A Summary of Case-control Research. *Lancet*, 341: 1432-1437.

Wacholder S (1994). When Measurement Errors Correlates with Truth: Surprising Effects of Nondifferential Misclassfication. *Epidemiology*, 6(2): 157-161.

Walter SD, Irwig LM (1988). Estimation of Test Error Rates, Disease Prevalence and Relative Risk from Misclassified Data: A Review. *Journal of Clinical Epidemiology*, 41(9): 923-937.

Watson RR (1988). Caffeine: Is It Dangerous to Health? *American Journal of Public Health*, 2(4): 13-22.

Wilks D (1996). President of the Coffee Association of Canada, personal communication.

Willet W (1990). Nutritional Epidemiology. Monographs in Epidemiology and Biostats, v.5. Chapter 12. Correction for the Effects of Measurement Error.

Willett W (1989). An Overview of Issues Related to the Correction of Non-Differential Exposure Measurement Error in Epidemiologic Studies. *Statistics in Medicine*, 8: 1031-1040.

Wu ML. Whittemore AS, Jung DK (1988). Errors in Reported Dietary Intakes. II. Long Term Recall. *American Journal of Epidemiology*, 128(5): 1137-1145.

Wynder EL, Hall NEL, Polansky M (1983). Epidemiology of Coffee and Pancreatic Cancer. *Cancer Research*, 43, 3900-3906.

Yesair DW, Branfman AR, Callahan MM (1985). Human Disposition and Some Biochemical Aspects of Methylxanthines. In Methylxanthine Beverages and Foods: Chemistry, Consumption, and Health Effects. Edited by Spiller GA, 215-234.

Zoumas BL, Krieser WR, Martin RA (1980). Theobromine and Caffeine Content of Chocolate Products. *Journal of Food Science*, 45, 314-316.

APPENDIX A: Scripts

Telephone Script
Initial Contact calls for Methylxanthine Study

Hi. My name is Don Maclean/ Joe Downey. I'm calling from the Ontario Cancer Treatment and Research Foundation. We are working on a research project related to cancer in Ontario. Your telephone number has been randomly selected from the phone book. We are looking for people who would be willing to answer a questionnaire sent by mail. The questionnaire takes 15-20 minutes to answer. Would you be willing to participate?

YES
Are you between the ages of 30-75? (insert the age range for the gender of the person on the phone)

NO
- thank them and hangup
- code reason for non-participation

YES
What is your exact age? (if there is pause add " I am filling out age quotas, that is why I ask.").

If they don't want to answer, say "all I need to know is if you between the ages of 30-44, 45-59 or 60-75"

Your phone number was randomly chosen from the phone book. The address I have is: (check address, ask if there is an appartment and get postal code). I will be sending this to you in the mail tomorrow. Thank you.

NO
We are looking for men between the ages of ___ and ___ and women between the ages of ___ and ____. Is there anyone in your household who qualifies?

If no, thank them and hang-up.

If yes,
Is that person home? May I speak to him/her?

If yes, go to the beginning script but skip asking them if they are age eligible

If no,
Do you think he/she would be willing to participate? If yes, take down information (check address, ask if there is an apartment number and get postal code). Would you mind telling him that I will be sending him this questionnaire. Great. Thank you.

May 29, 1995

Telephone Script

Reminder calls for Methylxanthine Study

Hello. May I please speak to (full name or Mr/Mrs so and so)?

If not available:
When is a good time to reach him/her? Thank you. I'll call him/her back later.

If available:
My name is Janet Brown from the Ontario Cancer Treatment and Research Foundation. About 12 days ago, I mailed you a research questionnaire, and I'm calling to see if you received it.

If NO,
I'm sorry that yours didn't reach you. Let me check that I have your correct address and I'll send you another one right away. I'd really appreciate if you could take a few minutes to help us with this study by completing the questionnaire when you receive it and returning it to us. Thank you very much for your time.

If YES,
Did you have a chance to read through it/ fill it out yet?

If YES,
Thank you very much for taking the time to fill it out. I look forward to receiving it.

If NO,
Do you have any questions or problems with the questionnaire?

If YES,
[Go through the questionnaire with the person.]

If NO,
We would appreciate if you could find the time to complete it. Do you think you might have the time to fill it out and send it to us in the next few days?

If NO,
May I ask why? (see objections)

QUESTIONS ABOUT THE STUDY

What organisation?
The Ontario Cancer Treatment and Research Foundation based in Toronto. We work in cooperation with the Princess Margaret Hospital.

What is the study for?
It's a population survey measuring the prevalence of different risk factors to cancer.

What are the questions about?
We are mainly interested in what you eat and drink, your exposure to tobacco, and your use of a few medications.

How did you get my name?
Your name was selected at random from the _____ (York, Durham, Halton-Peel, Hamilton-Wentworth) telephone book.

OBJECTIONS

Incapable of filling it out
- Handicap/Language
Maybe you could get someone there to help you with it. Is there someone in the family or a friend who could help you fill out the questionnaire?

-Difficulty understanding questionnaire
Were there any particular questions that you had problems with/didn't understand?

-If hadn't looked at the questionnaire yet
Most people find the questions quite easy to answer once they've had a chance to look at them.

Personal Information/Confidentiality
-- People's names are kept separate from the questionnaire. However, we have to have a way to get in touch with people for the purpose of clarifying answers.
-- I can certainly understand your concern. That's why all the answers on the questionnaire are strictly confidential and will only be used for research purposes. People's name are kept separate from the questionnaire and the results are reported in such a way that identification of individuals will never be possible.

[If still objects]. You may omit any questions you really don't want to answer but please fill out as much as you can. Is that all right with you?

This study is being conducted by Nancy Kreiger at the Ontario Cancer Treatment Research Foundation. If you have any concerns you may call her at 971-9800 ext. 1239.

Too busy
The questionnaire takes only 15-20 minutes to complete. Your participation in this study would greatly assist our research and contribute to the understanding of cancer. Do you think you will be able to help us? (by completing the questionnaire)

Insists still too busy
You don't have to send it in right away but if you do have some time over the next couple of weeks we would really appreciate your participation in this study

No interested
The questionnaire takes only 15-20 minutes to complete. Your participation in this study would greatly assist our research and contribute to the understanding of cancer. It is important that we get the information from everyone in the sample otherwise the results won't be very useful

No interested in receiving second questionnaire
Many people have found the questionnaire interesting. Let me send you another one, take a look at it and if you decide to complete it we would really appreciate it.

Wants to speak to the person responsible for the study
Janet Brown is the Project Coordinator. She can be reached at 971-9800.

APPENDIX B: Questionnaire

1 male
2 female

2. When were you born?

_____ / _____ / _____

day month year

3. How tall are you?

_____feet_____inches OR _____centimetres

4. How much did you weigh three (3) years ago?

_____pounds OR _____ kilograms

5. How much did you weigh when you were twenty-five (25) years old?

_____ pounds OR _____ kilograms

6. Were you living in Ontario at any time since April 1995?

1 yes
2 no
3 don't know

7. What is the hignest level of education that you have completed?

1 some or all elementary school (Grades 1-8)
2 some or all secondary school (Grades 9-12)
3 some or all post-secondary or university / college
4 don't know

8. Have you ever been diagnosed with any of the following conditions?

Please circle the appropriate answer in the centre part of the table, and write in the date, if applicable.

Medical Conditions		Date of Diagnosis
Pancreatitis	1 yes 2 no 3 don't know	_____ _____ month year
Insulin Dependent Diabetes	1 yes 2 no 3 don't know	_____ _____ month year
Non-Insulin Dependent Diabetes (controlled by diet alone)	1 yes 2 no 3 don't know	_____ _____ month year
Asthma	1 yes 2 no 3 don't know	_____ _____ month year
Gallbladder Disease	1 yes 2 no 3 don't know	_____ _____ month year
Gastrectomy (full or partial)	1 yes 2 no 3 don't know	_____ _____ month year

If 'yes', then please fill out the table below, otherwise, go to the next question.

Tobacco Product	Age Started	Usual number per day	Do/did you inhale into your chest?	Are you now using?	Age stopped
Filter cigarettes			1 yes 2 no	1 yes 2 no	
Non-filter cigarettes			1 yes 2 no	1 yes 2 no	
Other tobacco products			1 yes 2 no	1 yes 2 no	

If there is anything you would like to add about your 'smoking history', please do so below.

Next, we would like to know about your exposure to environmental tobacco smoke, also known as second-hand smoke.

10. When you were a child, how many hours per day (on average) were you exposed to the tobacco smoke of other people?

0 not exposed
1 less than 3 hours per day
2 3-8 hours per day
3 9 or more hours per day
4 don't know

11. About 3 years ago, how many hours per day (on average) were you exposed to the tobacco smoke of other people? Please include working and non-working days.

0 not exposed
1 less than 3 hours per day
2 3-8 hours per day
3 9 or more hours per day
4 don't know

The following questions deal with your use of various medications.

12. Please circle the name(s) of the **breathing** medication(s) (listed below) which you have ever used for more than 15 days in any month, or 3 times a week for three months or more.

If you did not use any, please check here ☐:

PMS Theophylline	Somophyllin	Theolair
Quibron	Theo-dur	Theophylline by Injection
Slo-bid	Theochron	Uniphyl
		Other_____

Please complete the chart below. Add others on the back page of the questionnaire if necessary.

Name of **Breathing** Medication	Usual number of pills or injections per day	Years Used	Usual number of months of use per year
		From_____ To _____	
		From_____ To _____	
		From_____ To _____	

Atasol	Fiorinal	282
Acetaminophen with Codeine	Instantine	292
C2	Kalmex	Tylenol with Codeine(1,2,or3)
Cafergot	Midol	Wake-ups
Darvon N(compound pulvule 405)	Novopropoxyn	Other(s)_____
Ergodryl	692	_____
	Stay Alert	

Please complete the chart below for all medications that you circled. Add others on the back page of the questionnaire if necessary.

Name of Medication (listed above)	Usual number of pills or injections per day	Years Used	Usual number of months of use per year
		From_____ To _____	
		From_____ To _____	
		From_____ To _____	

If you would like to add anything about your medical history, please do so on the back page of the questionnaire.

Now we would like to find out about your pattern of non-alcoholic beverage consumption.

14. Did you ever drink coffee, tea, chocolate milk or soft drinks? *Please circle your answer before moving on.*

1 yes
2 no
3 don't know

If 'yes', then please fill out the table below, otherwise, go to the next question.

Beverage Type	Age Started	Usual Number Servings Per Day	Are you now consuming?	Age stopped
Regular coffee			1 yes 2 no	
Instant coffee			1 yes 2 no	
Decaffeinated coffee			1 yes 2 no	
Espresso, cappuccino			1 yes 2 no	
Regular tea			1 yes 2 no	
Herbal tea			1 yes 2 no	
Cocoa, hot chocolate			1 yes 2 no	
Chocolate milk			1 yes 2 no	
Cola soft drinks (e.g., Coke, Pepsi)			1 yes 2 no	
Non-cola soft drinks (e.g., 7-Up, rootbeer)			1 yes 2 no	

Your Servings of Alcoholic Beverages, About 3 Years Ago								
Alcoholic Beverage Type	Serving Size	None	Less than 1 per week	1-6 per week	1 per day	2-7 per day	8 or more per day	don't know
Beer	12 ounces							
Spirits / liquor (gin, rye, vodka, brandy)	1 ounce							
Red wine	4 ounces							
White wine	4 ounces							
Fortified wines (port, sherry, vermouth)	1 ounce							

16. About 3 years ago, how many servings of the following foods did you usually eat?

Please check one box for each food item, including those you did not use.

Your Consumption, About 3 Years Ago								
Food	Serving Size	None	Less than 1 per week	1-6 per week	1 per day	2-7 per day	8 or more per day	don't know
Pasta, rice, potatoes	4 ounces							
Eggs	1 egg							
White bread	1 slice							
Cereal	1 cup							
Chocolate bar(s)	1							

17. About 3 years ago, how many servings of the following foods did you usually eat?

Please check one box for each food item, including those you did not use.

Your Fruit and Vegetable Consumption, About 3 Years Ago								
Fruits / Vegetables	Serving Size	None	Less than 1 per week	1-6 per week	1 per day	2-7 per day	8 or more per day	don't know
Carrots	4 ounces							
Spinach, chard, mustard greens	4 ounces							
Cabbage, broccoli, Brussels sprouts, cauliflower	4 ounces							
Other green vegetables	4 ounces							
Citrus fruits and juices	4 ounces 8 oz. juice							
Other fruits and juices	4 ounces 8 oz. juice							

Your Servings of Alcoholic Beverages, About 15 Years Ago								
Alcoholic Beverage Type	Serving Size	None	Less than 1 per week	1-6 per week	1 per day	2-7 per day	8 or more per day	don't know
Beer	12 ounces							
Spirits / liquor (gin, rye, vodka, brandy)	1 ounce							
Red wine	4 ounces							
White wine	4 ounces							
Fortified wines (port, sherry, vermouth)	1 ounce							

19. About 15 years ago, how many servings of the following foods did you usually eat?

Please check one box for each food item, including those you did not use.

Your Consumption, About 15 Years Ago								
Food	Serving Size	None	Less than 1 per week	1-6 per week	1 per day	2-7 per day	8 or more per day	don't know
Pasta, rice, potatoes	4 ounces							
Eggs	1 egg							
Cheese	1 in. cube							
White bread	1 slice							
Cereal	1 cup							
Chocolate Bar(s)	1							

20. About 15 years ago, how many servings of the following foods did you usually eat?

Please check one box for each food item, including those you did not use.

Your Fruit and Vegetable Consumption, About 15 Years Ago								
Fruits / Vegetables	Serving Size	None	Less than 1 per week	1-6 per week	1 per day	2-7 per day	8 or more per day	don't know
Carrots	4 ounces							
Spinach, chard, mustard greens	4 ounces							
Cabbage, broccoli, Brussels sprouts, cauliflower	4 ounces							
Other green vegetables	4 ounces							
Citrus fruits and juices	4 ounces 8 oz. juice							
Other fruits and juices	4 ounces 8 oz. juice							

If yes, please answer parts a,b,c, and d.

21a How many times have you been pregnant? (Include live births, miscarriages, still births, and abortions.)
_____ pregnancy(ies)

21b How old were you when you were first pregnant? _____ years old

21c How many of your pregnancies were live births? _____ live births

21d During any of your pregnancies were you given any medication to prevent bleeding or miscarriage?
1 yes
2 no
3 don't know

22. Prior to 3 years ago, had you ever taken oral contraceptives (birth control pills), for any reason, for 6 months or more?
1 yes
2 no
3 don't know

23. Prior to 3 years ago, had you ever taken female hormones (estrogens such as Premarin), for any reason, for 6 months or more?
1 yes
2 no
3 don't know

Everyone please continue here.

24. Did someone help you fill out this questionnaire?
1 yes
2 no

If yes, how is this person is related to you? _____

25. If we do not understand some information you provided in the questionnaire, we would like to be able to contact you by telephone. If this is acceptable to you, please provide your telephone number below.

(_____) _____-_____
area code telephone number

26. Finally, because accurate data are so important, please check through all the questions and tables, to be sure nothing was missed.

Is there anything else you would like to tell us? If so, please use the bottom of this page, or add paper if necessary.

Your contribution to this study is greatly appreciated. If you would like a summary of the results, please print your name and address on the back of the return envelope, and we will send the results after the analysis is completed.

Thank you.

APPENDIX C: Explanatory letter

May 29, 1997

FIELD(1)
FIELD(2)
FIELD(3)
FIELD(4)

Dear FIELD(1):

Thank you for agreeing to help in our research study. As you know, we are interested in carrying out research into the causes of cancer, which would be a comparison between people who have had cancer and people who have not. Before we can do this, however, we need to collect some preliminary information for a larger study we will be conducting. Thus, your participation is vital in helping us design a better study.

As discussed on the telephone, the research questionnaire is enclosed. It is important that every questionnaire be completed, although your participation is completely voluntary. It should take about 15-20 minutes to complete. Please complete the questionnaire and return it in the postage-paid envelope as soon as possible. It is important that you complete the questionnaire yourself (with assistance, if necessary), and not pass it on to another person.

All responses will be treated as confidential and will be used only for research purposes. The information you provide will be analyzed and reported in terms of groups only. A number appears on your questionnaire; this is necessary for administrative reasons, so that we know not to contact you once you have returned the completed questionnaire.

If you have any questions about this study, or have difficulty filling out the questionnaire, please contact me at the above address or call me or the study co-ordinator, Janet Brown, at (416) 971-9800 ext. 1220 (collect if long distance).

Thank you for your co-operation and assistance.

Sincerely,

Nancy Kreiger, M.P.H., Ph.D.
Principal Investigator

APPENDIX D: Kappa formulae

Kappa formulae

Let n_{ij} denote the number of observations in the *ith* row and the *jth* column of a two-way table of caffeine intake from regular coffee by total caffeine intake. Then n_{ij} is the number of pairs in the *ijth* cell of the table. Let n_{i+} and n_{+j} represent the *ith* row and *jth* column marginal total respectively, n_{++} the total sample size, and w_{ij} the weight associated with the ijth cell. The weighted kappa statistic takes the form

$$k_w = (p_{ow} - p_{ew})/(1 - p_{ew})$$

where $p_{ow} = 1/n_{++} \Sigma\Sigma w_{ij}$,
the observed weighted proportion of pairs in agreement, and
$p_{ew} - (1/n_{++})^2 \Sigma\Sigma w_{ij} n_{i+} n_{+j}$
the weighted proportion of pairs in agreement expected under a model of statistical independence.

There are many weighting schemes possible; one being the squared error weights

$w_{ij} = 1 - \{(i-j)^2/(r-1)^2\}$ where *r* is the number of categories and *i* and *j* are category ranks.

Values of weighted kappa differ depending on the weighting scheme used. The squared error weights have been suggested as being the most easily interpretable, since under this weighting scheme, weighted kappa is asymptotically equivalent to the intraclass correlation computed using category ranks to score responses (Fleiss & Cohen, 1973); it is therefore the scheme we have chosen for this project.

Unweighted kappa is a special case of the weighted kappa where
$W_{ij} = 1$ for $i=j$ and $w_{ij} = 0$ otherwise

Fleiss (1981) suggests that values of kappa (weighted or unweighted) $\geq .75$ signify excellent agreement; values ≤ 0.40 signify poor agreement. To test the hypothesis that $k > 0$, i.e., agreement beyond that expected by chance, standard error will be calculated as follows:

$$Var(k_w) = \Sigma\Sigma p_{i+} p$$

$$Var(k_w) = \Sigma\Sigma p_{i+} p_{+j} \{w_{ij} - (w_{i+} + w_{+j})\}^2 - p_{ew}$$
$$n_{++}(1-p_{ew})^2$$

where $w_{i+} = 1/n_{++} \Sigma n_{+j} w_{ij}$
$w_{+j} = 1/n_{++} \Sigma n_{j+} w_{ij}$

APPENDIX E: Correction formulae

Correction formulae

Derivations/simplification/ from formulae presented in Kleinbaum, Kupper & Morgenstern (1982) assuming non-differential misclassification/bias of disease and given that the specificity in this study is 1.0.

Correcting for misclassification in study data

a. Observed data

	E	~E	
D	a	b	N_D
~D	c	d	$N_{\sim D}$

b. Corrected data

	E	~E
D	A"	B"
~D	C"	D"

$$OR(observed) = \frac{a*d}{b*c} = x \qquad OR(corrected) = \frac{A"*D"}{B"*C"}$$

where a, b, c, d are the numbers that are reported in the article and A", B", C", D" are the corrected numbers based on the formulas given below.

ϕ_E = sensitivity given exposure = based on misclassification estimates of this study = probability that a person who is exposed will be classified as exposed = $C_{0,0} / (C_{0,0}+C_{1,0})$.

	True caffeine exposure	
Classified exp.	E_0	E_1
E_o	$C_{0,0}$	$C_{0,1}$
E_1	$C_{1,0}$*	$C_{1,1}$

$A" = (N_D - b)/ \phi_E$
$B" = (N_D * \phi_E - a)/ \phi_E$
$C" = (N_{\sim D} - d)/ \phi_E$
$D" = (N_{\sim D} * \phi_E - c)/ \phi_E$

APPENDIX F: Misclassification matrices

Misclassification matrices

DAILY INTAKE: males 30-44

Level of exposure

Coffee \ Total caffeine	0	1-115	116-230	231+	
0	2	8	2	3	15 (17.7%)
1-115	0	1	7	4	12 (14.2%)
116-230	0	0	6	20	26 (30.6%)
231+	0	0	0	32	32 (37.7%)
	2 (2.4%)	9 (10.6%)	15 (17.7%)	59 (69.4%)	85 (100.0%)

DAILY INTAKE: males 45-59

Level of exposure

Coffee \ Total caffeine	0	1-115	116-230	231+	
0	1	4	2	4	11 (15.1%)
1-115	0	2	6	3	11 (15.1%)
116-230	0	0	2	15	17 (23.3%)
231+	0	0	0	34	34 (46.6%)
	1 (1.4%)	6 (8.2%)	10 (13.7%)	56 (76.7%)	73 (100.0%)

DAILY INTAKE: males 60-75

Level of exposure

Coffee \ Total caffeine	0	1-115	116-230	231+	
0	1	2	3	1	7 (11.9%)
1-115	0	2	4	3	9 (15.3%)
116-230	0	0	4	13	17 (28.8%)
231+	0	0	0	26	26 (44.1%)
	1 (1.7%)	4 (6.8%)	11 (18.6%)	43 (72.9%)	59 (100.0%)

DAILY INTAKE: females 30-44

Level of exposure

Coffee \ Total caffeine	0	1-115	116-230	231+	
0	2	9	9	1	21 (26.3%)
1-115	0	4	8	3	15 (18.8%)
116-230	0	0	3	19	22 (27.5%)
231+	0	0	0	22	22 (27.5%)
	2	13	20	45	80 (100.0%)
	(2.5%)	(16.3%)	(25.0%)	(56.3%)	

DAILY INTAKE: females 45-59

Level of exposure

Coffee \ Total caffeine	0	1-115	116-230	231+	
0	0	8	3	4	15 (16.7%)
1-115	0	6	16	3	25 (27.8%)
116-230	0	0	2	19	21 (23.3%)
231+	0	0	0	29	29 (32.2%)
	0	14	21	55	90 (100.0%)
	(0.0%)	(15.6%)	(23.3%)	(61.1%)	

DAILY INTAKE: females 60-75

Level of exposure

Coffee \ Total caffeine	0	1-115	116-230	231+	
0	3	11	3	4	21 (22.3%)
1-115	0	8	11	7	26 (27.7%)
116-230	0	0	2	19	21 (22.3%)
231+	0	0	0	26	26 (27.7%)
	3	19	16	56	94 (100.0%)
	(3.9%)	(20.2%)	(17.0%)	(59.6%)	

Level of exposure

Coffee \ Total caffeine	0-999	1000-3199	3200-5999	6000	
0-999	4	8	7	1	20 (23.5%)
1000-3199	0	7	8	6	21 (24.7%)
3200-5999	0	0	9	16	25 (29.4%)
6000	0	0	0	19	19 (22.4%)
	4 (4.7%)	15 (17.7%)	24 (28.2%)	42 (49.4%)	85 (100.0%)

LIFETIME INTAKE: males 45-59

Level of exposure

Coffee \ Total caffeine	0-199	200-6699	6700-11699	11700+	
0-199	1	5	5	6	17 (24.3%)
200-6699	0	2	14	2	18 (25.7%)
6700-11699	0	0	9	8	17 (24.3%)
11700+	0	0	0	18	18 (25.7%)
	1 (1.4%)	7 (10.0%)	28 (40.0%)	34 (48.6%)	70 (100.0%)

LIFETIME INTAKE: males 60-75

Level of exposure

Coffee \ Total caffeine	0-5999	6000-10999	11000-16999	17000+	
0-5999	3	6	4	2	15 (25.2%)
6000-10999	0	2	4	8	14 (23.7%)
11000-16999	0	0	4	4	18 (30.5%)
17000+	0	0	0	12	12 (20.3%)
	3 (5.1%)	8 (13.6%)	12 (20.3%)	36 (61.0%)	59 (100.0%)

Level of exposure

Coffee \ Total caffeine	0	1-1799	1800-3999	4000+	
0	0	7	15	3	25 (32.1%)
1-1799	0	5	4	5	14 (18.0%)
1800-3999	0	0	16	13	19 (24.4%)
4000+	0	0	0	20	20 (25.6%)
	0 (0.0%)	12 (15.4%)	25 (32.1%)	41 (52.6%)	78 (100.0%)

LIFETIME INTAKE: females 45-59

Level of exposure

Coffee \ Total caffeine	0-999	1000-3999	4000-7999	8000+	
0-999	1	6	7	7	21 (23.9%)
1000-3999	0	2	12	9	23 (26.1%)
4000-7999	0	0	6	17	23 (26.1%)
8000+	0	0	0	21	21 (23.9%)
	1 (1.4%)	8 (9.1%)	25 (28.4%)	54 (61.4%)	88 (100.0%)

LIFETIME INTAKE: females 60-75

Level of exposure

Coffee \ Total caffeine	0	1-5799	5800-12399	12400+	
0	1	6	4	1	22 (25.0%)
1-5799	0	3	12	7	22 (25.0%)
5800-12399	0	0	8	14	22 (25.0%)
12400+	0	0	0	22	22 (25.0%)
	1 (1.1%)	9 (10.2%)	24 (27.3%)	54 (61.4%)	88 (100.0%)

APPENDIX G: Kappa worksheets

unweighted kappa

n

	1	2	3	4	5	6	
1	7	0	0	0	0	0	7
2	36	0	0	0	0	0	36
3	25	4	0	0	0	0	29
4	46	11	34	2	0	0	93
5	30	4	43	49	0	0	126
6	13	1	15	45	62	51	187
	157	20	92	96	62	51	478

weight

	1	2	3	4	5	6
1	1	0	0	0	0	0
2	0	1	0	0	0	0
3	0	0	1	0	0	0
4	0	0	0	1	0	0
5	0	0	0	0	1	0
6	0	0	0	0	0	1

Pow

	1	2	3	4	5	6	
1	7	0	0	0	0	0	
2	0	0	0	0	0	0	
3	0	0	0	0	0	0	
4	0	0	0	2	0	0	
5	0	0	0	0	0	0	
6	0	0	0	0	0	51	
							60

pow 0.125523

Pew

	1	2	3	4	5	6	
1	1099	0	0	0	0	0	
2	0	720	0	0	0	0	
3	0	0	2668	0	0	0	
4	0	0	0	8928	0	0	
5	0	0	0	0	7812	0	
6	0	0	0	0	0	9537	
							30764

Pew 0.134644

kw -0.01054

var		1	2	3	4	5	6	
	1	0.002076	9.99E-05	0.000427	0.000805	0.000666	0.000809	
	2	7.89E-05	0.002456	0.000152	0.000845	0.000911	0.001507	
	3	0.000855	0.000182	0.006513	0.001825	0.001637	0.002205	
	4	0.002967	0.000621	0.002561	0.014284	0.013784	0.007276	
	5	0.001804	0.000464	0.001839	0.005567	0.012585	0.007632	
	6	0.001892	0.000542	0.002109	0.007131	0.006958	0.010523	
								0.124587

VAR(kw)	0.000297
SE (kw)	0.017246

	1	2	3	4	5	6
wi+	0.328452	0.041841	0.192469	0.200837	0.129707	0.106695
wj+	0.014644	0.075314	0.060669	0.194561	0.263598	0.391213

* kap_tab.xls

Worksheet for kappa statistics for calculations of daily regular coffee and total daily caffeine intake for the whole sample using squarred weights

		Regular coffee intake						
n		1	2	3	4	5	6	Total
	1	7	0	0	0	0	0	7
	2	36	0	0	0	0	0	36
	3	25	4	0	0	0	0	29
	4	46	11	34	2	0	0	93
	5	30	4	43	49	0	0	126
	6	13	1	15	45	62	51	187
Total		157	20	92	96	62	51	478

weight		1	2	3	4	5	6	
	1	1.00	0.96	0.84	0.64	0.36	0.00	3.80
	2	0.96	1.00	0.96	0.84	0.64	0.36	4.76
	3	0.84	0.96	1.00	0.96	0.84	0.64	5.24
	4	0.64	0.84	0.96	1.00	0.96	0.84	5.24
	5	0.36	0.64	0.84	0.96	1.00	0.96	4.76
	6	0.00	0.36	0.64	0.84	0.96	1.00	3.80

Pow		1	2	3	4	5	6	
	1	7.00	0.00	0.00	0.00	0.00	0.00	
	2	34.56	0.00	0.00	0.00	0.00	0.00	
	3	21.00	3.84	0.00	0.00	0.00	0.00	
	4	29.44	9.24	32.64	2.00	0.00	0.00	
	5	10.80	2.56	36.12	47.04	0.00	0.00	
	6	0.00	0.36	9.60	37.80	59.52	51.00	
								394.52

pow 0.825356

Pew		1	2	3	4	5	6	
	1	1099.00	134.40	540.96	430.08	156.24	0.00	
	2	5425.92	720.00	3179.52	2903.04	1428.48	660.96	
	3	3824.52	556.80	2668.00	2672.64	1510.32	946.56	
	4	9344.64	1562.40	8213.76	8928.00	5535.36	3984.12	
	5	7121.52	1612.80	9737.28	11612.16	7812.00	6168.96	
	6	0.00	1346.40	11010.56	15079.68	11130.24	9537.00	
								158594

Pew 0.694116

kw **0.429051**

Worksheet for calculation of standard error of kappa of daily regular coffee and total daily caffeine intake for the whole sample using squarred weights

	1	2	3	4	5	6
wi+	0.71	0.83	0.88	0.85	0.73	0.52
wj+	0.36	0.62	0.80	0.91	0.93	0.87

var	1	2	3	4	5	6
1	1.9E-05	8.2E-05	1.3E-03	2.8E-03	3.1E-03	3.9E-03
2	1.3E-03	6.5E-04	6.6E-03	1.2E-02	1.2E-02	1.5E-02
3	3.1E-03	7.4E-04	5.4E-03	8.3E-03	7.4E-03	8.0E-03
4	2.0E-02	3.2E-03	1.8E-02	2.2E-02	4.2E-02	1.6E-02
5	4.6E-02	5.6E-03	2.5E-02	2.4E-02	1.5E-02	1.2E-02
6	1.0E-01	1.0E-02	3.6E-02	2.7E-02	1.2E-02	6.6E-03

0.5323

VAR(kw) 0.001129
SE (kw) 0.033608

APPENDIX H: Correction worksheets

MacMahon (1981)

Actual population (mislcassified)

	<1	1_2	3_4	5+	
D	20.0	153.0	106.0	88.0	367
~D	32.0	119.0	74.0	82.0	307

Ss(1,2-4)	0.6904
Ss(1-2,3-4)	0.3599
Ss(1-3,4)	0.2037

True population

	<1	1_2	3_4	5+	
D	-135.6	-36.4	107.1	432.0	367
~D	-91.3	-35.1	30.9	402.5	307

	OR(1,1)	OR(2,1)	OR(3,1)	OR(4,1)
ACTUAL	1.00	2.06	2.29	1.72
TRUE	1.00	0.70	2.33	0.72
Bias		0.66	-0.02	0.58

	TRUE				
classified	<1	1_2	3_4	5+	
0-1	45	86	44	5	180
1_2	0	36	128	24	188
3_4	0	0	34	57	91
5+	0	0	0	22	22
	45	122	206	108	481

Comparibility coffee

	<1	1_2	3_4	5+	
Brown	37.4	39.1	18.9	4.6	100
MacMahon controls	10.4	38.8	24.1	26.7	100

Jain, Howe, St.Louis and Miller (991)

Actual population (mislcassified)

	0	0-39	40-110	110+	
D	25.0	69.0	76.0	76.0	246
~D	40.0	136.0	174.0	154.0	504

Ss(1,2-4)	0.7689
Ss(1-2,3-4)	0.6310
Ss(1-3,4)	0.4899

True population

	0	0-39	40-110	110+	
D	-41.4	46.5	85.7	155.1	246
~D	-99.5	83.7	205.4	314.4	504

	OR(1,1)	OR(2,1)	OR(3,1)	OR(4,1)
ACTUAL	1.00	0.81	0.70	0.79
TRUE	1.00	1.34	1.00	1.18
	Bias	-0.65	-0.43	-0.50

	TRUE				
classified	0	0-39	40-110	110+	
0	5	48	38	24	115
0-39	0	35	71	12	118
40-110	0	0	86	65	151
110+	0	0	0	97	97
	5	83	195	198	481

COLLAPSED CATEGORIES

Actual population

	0-110	110+
D	170.0	76.0
~D	350.0	154.0

OR = 1.02

True population

	0-110	110+
D	90.9	155.1
~D	189.6	314.4

OR = 1.03

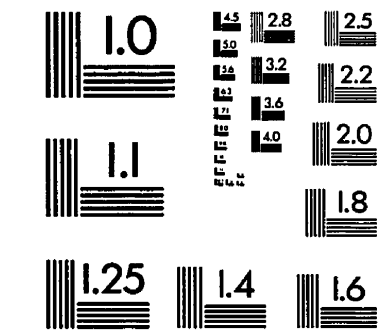
Comparability cup years of coffee

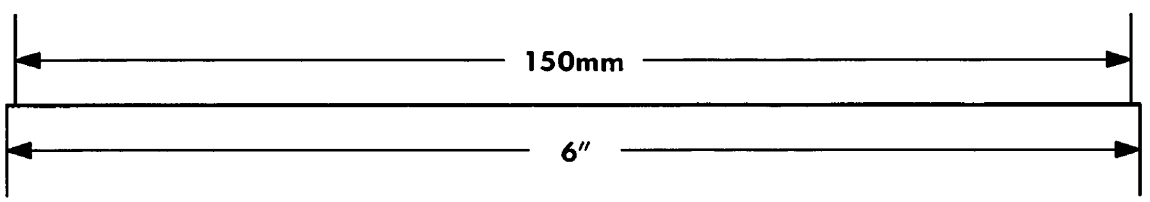
	0	0-39	40-110	110+	
Brown	23.9	24.5	31.4	20.2	100.0
Jain et al.	7.9	27.0	34.5	30.6	100.0

Data mtx.quest13 daily intake of regular coffee vs. caffeine (in quartiles)							
		ALL					
True population							
	0	1-115	116-230	231+			
D	0.6	7.9	15.1	76.4	100		
~D	1.5	13.6	19.5	65.5	100		
Actual population (misclassified)							
	0	1-115	116-230	231+			
D	32.2	20.7	19.5	27.6	100	Ss(0,1-3)	0.6815
~D	32.8	23.4	20.1	23.6	100.0	Ss(0-1,2-3)	0.5148
						Ss(0-2,3)	0.3610
	OR(0,0)	OR(1,0)	OR(2,0)	OR(3,0)			
TRUE	1.00	1.50	2.00	3.00			
actual	1.00	0.90	0.99	1.19			
	Bias	0.40	0.50	0.60			
	TRUE						
classified	0	1-115	116-230	231+			
0	7	61	46	43	157		
1-115	0	4	45	63	112		
116-230	0	0	2	94	96		
231+	0	0	0	113	113		
	7	65	93	313	478		

Data mtx.quest13 lifetime inakte of regular coffee vs. caffeine								
		ALL						
True population								
	0	1-4149	4150-8999	9000+				
D	0.1	11.5	23.2	65.3	100			
~D	0.2	18.6	28.2	53.0	100			
Actual population (misclassified)								
	0	1-4149	4150-8999	9000+				
D	21.7	23.8	23.1	31.3	100		Ss(0,1-3)	0.7837
~D	21.8	28.2	24.6	25.4	100.0		Ss(0-1,2-3)	0.6158
							Ss(0-2,3)	0.4798
	OR(0,0)	OR(1,0)	OR(2,0)	OR(3,0)				
TRUE	1.00	1.50	2.00	3.00				
actual	1.00	0.85	0.95	1.24				
	Bias	0.43	0.53	0.59				
	TRUE							
classified	0	1-4149	4150-8999	9000+				
0	1	45	25	31	102			
1-4149	0	42	66	24	132			
4150-8999	0	0	41	74	115			
9000+	0	0	0	119	119			
	1	87	132	248	468			

Data mtx.quest13 daily intake of regular coffee vs. caffeine (in quartiles)							
		Males 30-44					
True population							
	0	1-115	116-230	231+			
D	0.9	6.1	13.5	79.5	100		
~D	2.4	10.6	17.7	69.4	100		
Actual population (misclassified)							
	0	1-115	116-230	231+			
D	23.6	23.6	20.4	32.4	100	Ss(0,1-3)	0.7711
~D	24.7	25.9	21.2	28.2	100	Ss(0-1,2-3)	0.5676
						Ss(0-2,3)	0.4068
	OR(0,0)	OR(1,0)	OR(2,0)	OR(3,0)			
TRUE	1.00	1.50	2.00	3.00			
actual	1.00	0.96	1.01	1.20			
	Bias	0.36	0.49	0.60			
	TRUE						
classified	0	1-115	116-230	231+			
0	2	9	4	6	21		
1-115	0	0	10	12	22		
116-230	0	0	1	17	18		
231+	0	0	0	24	24		
	2	9	15	59	85		





APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989