INVESTIGATION INTO THE RELATIONSHIP BETWEEN PHYSICAL ACTIVITY AND TOTAL PLASMA HOMOCYSTEINE

by

Amy Schneeberg

A thesis submitted to the Department of Community Health and Epidemiology in conformity with the requirements for the degree of Master of Science

Queen's University

Kingston, Ontario, Canada

September, 2007

Abstract

Background: A beneficial effect of physical activity on the risk of cancer at several sites has been consistently observed. Biologic mechanisms that may underlie this relationship are not well understood. A potential mechanism explaining this relationship for some cancer sites is the influence of physical activity on methionine-homocysteine biosynthesis. High levels of total plasma homocysteine concentration (tHcy) indicate a breakdown in this biochemical process. This cycle's influences on DNA methylation and endogeneous agents involved in oxidative stress are potential mechanisms linking methionine-homocysteine biosynthesis to cancer risk. This research is nested within a larger cross sectional study of healthy volunteers recruited from centers in Ontario and Nova Scotia aimed at understanding modifiable risk factors for cancer.

Purpose: This research sought to elucidate the relationship between physical activity and tHcy level.

Methods: The target population was healthy male and female subjects aged 20-50. Participants donate a 12ml blood sample after an overnight fast for analysis of tHcy and dietary factors and complete a questionnaire including a physical activity profile for the past month (adapted from the International Physical Activity Questionnaire [IPAQ]) and established predictors of tHcy level such as coffee and alcohol consumption. Multiple linear regression is used to model the relationship between tHcy and physical activity measures while controlling for potential confounders.

Results: Analysis on 171 participants has been carried out. Mean tHcy for five quintiles of physical activity (from lowest physical activity score to highest) were found to be 8.40µmol/L (7.76-9.05), 8.60µmol/L (8.00, 9.22), 9.24µmol/L (8.66, 9.81), 8.23 µmol/L(7.64, 8.82), and 8.70µmol/L (8.09, 9.31).

Conclusions: The findings of this research do not support a relationship between physical activity and total plasma homocysteine concentration. Results of this study suggest that homocysteine is not a mediating factor for the relationship observed between physical activity and cancer.

Acknowledgements

I would like to sincerely thank Dr. Will King for his continued guidance and support throughout this project, and for being available to answer my hundreds of questions. Similarly, thanks to Dr. Ian Casson for being available for discussions, providing direction and suggestions and for his thoroughness and attention to detail. Also, thanks to Dr. Thomas Massey for his expertise and counsel on this project.

My kind thanks goes to Gwyneth Kearney for all of her work, her kind support, and for making the office brighter and always being there for a laugh or a coffee. As well to Janet Ashbury a kinder person I will never meet! Thank you for your patient listening and friendship.

To my office mates and follow students many thanks for the encouragement and comic/gossip relief when needed.

I would like to thank faculty, staff and students at the Department of Community Health and Epidemiology for their role in making this program and my master's experience so enjoyable.

To my family (Mom, Unc, and Kevin) for always being there to read pages and pages on homocysteine, for listening to my trials (over and over), for examining a poster for the umpteenth time, for all of your unconditional love and support, and for the smiles and good times THANK YOU!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BE..................Backward Elimination

BMI ....Body Mass Index

CI...................Confidence Interval

CV ....Coefficient of Variation

DNA ....Deoxyribonucleic Acid

GSH...............Glutathione

HEPA ....Health Enhancing Physical Activity

HDL...............High-Density Lipoprotein

IPAQ..............International Physical Activity Questionnaire

LDL ....Low-Density Lipoprotein

MET ....Metabolic Equivalent

OR ....Odds Ratio

SAH...............S-adenosyl-homocysteine

SAM ....S-adenosyl-methionine

tHcy. ....Total Plasma Homocysteine

THM..............Trihalomethanes

Std Dev..........Standard Deviation

SES................Socio- Economic Status

Chapter 1: Introduction

It is well documented that increased physical activity decreases the risk of several types of cancer. The biological mechanism that explains this relationship is not clear, and it is thought to be affected by a number of different factors. One hypothesis is an effect of physical activity on methionine-homocysteine biosynthesis. There is evidence a dysfunction in the methionine-homocysteine biosynthesis can increase cancer risk and it is known that elevated total plasma homocysteine (tHcy) is one indication of a dysfunction in this cycle. It is biologically plausible that there is an association between tHcy and physical activity, however there are gaps and inconsistencies in the literature pertaining to this relationship. In this thesis the relationship between physical activity and tHcy was examined with the hypothesis that a U shaped curve describes this relationship (low physical activity associated with high tHcy, moderate/health-enhancing physical activity associated with low tHcy, and very high levels of physical activity associated with high tHcy).

Cancer is the leading cause of death in Canadians (Statistics Canada, 2004), with a postulated increase in incidence in years to come (NCIC, 2007). Based on current incidence rates 39% of Canadian women and 44% of Canadian men will develop cancer in their lifetimes (NCIC, 2007). For this reason it is important to identify modifiable risk factors for this disease. As well, it is important to develop an understanding of the relationship between the risk modifiers and cancer as this knowledge can lead to a better understanding of the etiology of this disease. Understanding etiology is essential to the development of treatment and prevention programs.

Physical activity is one modifiable factor that has been found to reduce the risk of cancer at a number of sites. Over 180 observational epidemiologic studies have examined the relationship between physical activity and cancer. The evidence for a beneficial effect of physical activity on cancer risk ranges from possible to convincing depending on the site of the cancer. For colon cancer there is strong evidence that indicates a reduced risk with increased physical activity. A recent review found that of 56 studies conducted on colon and colorectal cancer, 48 demonstrated a reduction in cancer risk among the most physically active participants (Quadrilatero, 2003). A review of postulated mechanisms between physical activity and colon cancer suggests several potential mechanisms through which activity exerts its protective effects (Friedenreich, 2002). These mechanisms include physical activity's effect on: gastrointestinal transit time, immune function, prostaglandin levels, insulin, and bile acid metabolism. However, it is not clearly understood what truly mediates this relationship and it is likely to be a combination of different factors. This research investigated the relationship between physical activity and total plasma homocysteine (tHcy) as a potential mediating factor for this association.

This research pertains to tHcy not for its direct health effects (i.e. relationship between tHcy and cardiovascular disease), but as a biomarker of the methionine-homocysteine biosynthesis (or methionine metabolism). Elevated tHcy is one indicator that there has been a disruption in this cycle which can lead to two potential cancer causing mechanisms. It is biologically plausible that physical activity could affect methionine-metabolism and that this cycle could play a role in mediating the relationship between physical activity and cancer.

My role within this research started with the development of the study instrument. I was involved in choosing the physical activity questionnaire and implemented pilot testing of this instrument. I also helped design participant recruitment materials and was involved in data collection. I entered and cleaned the data for analysis and created interpretable variables to examine as potential confounders. Finally I completed an independent analysis of the data collected to date.

This thesis is organized in five sections. A review of the relevant literature for this project is presented in chapter 2. Chapter 3 gives an overview of the general methods, including study design, data sources, and analysis plan. The results of the data analysis are presented in Chapter 4 and a summary of the study's main findings, its limitations and possible future directions are discussed in Chapter 5.

Chapter 2.0: Literature Review

This research investigated the relationship between physical activity and total plasma homocysteine concentration (tHcy). Total plasma homocysteine is a biomarker of the homocysteine-methionine cycle (methionine metabolism) and dysfunction in this cycle can lead to several cancer-causing mechanisms including abnormal DNA methylation (see page 13 for conceptual model). The ultimate goal of this research was to provide a better understanding of the effects of physical activity on plasma homocysteine concentration and thus the homocysteine-methionine cycle in the context of cancer etiology.

2.1 Physical Activity

The accepted scientific definition of physical activity is 'any voluntary movement produced by skeletal muscles that results in energy expenditure' (Caspersen, 1985). Although all people take part in physical activity, the frequency and type varies from person to person. In research, physical activity is often measured by the intensity, frequency, and duration of activity performed during different segments of daily life, ie. At work, at home and at leisure (Caspersen, 1985).

2.1.1 Physical Activity Epidemiology

Physical activity has been the exposure of interest in numerous epidemiological studies and many relationships between physical activity and disease outcomes have been observed. Due to associations that have been observed and the biologic plausibility for many others that could provide prevention strategies and insight into disease etiology, physical activity epidemiology is a growing field.

2.1.2 Measurement of Physical Activity

Measuring participation in physical activity for research purposes presents several questions and challenges. The interpretation of 'physical activity' can vary from study to study so a clear definition needs to be identified before choosing a study instrument to quantify this variable. It needs to be determined if data on physical activity from all domains of life is to be gathered or if only specific domains are of interest (i.e., leisure-time, gardening/yard work, household chores, physical activity for transport, occupational). The time frame of interest needs to be defined, and, if necessary, seasonal and weekly variation in physical activity needs to be addressed. The method of defining different types of physical activity (i.e., by symptoms, e.g., breathlessness, and sweating or by specific activity or by quality e.g., moderate vs. vigorous) needs to be determined (US Department of Health and Human Services, 1996).

The measurement of physical activity has been approached in several manners in past epidemiological studies including calorimetry, job classification, survey procedures, physiological markers of physical activity, behavioural observation, mechanical and electronic monitoring, heart rate monitoring, motion sensors, and dietary measures (Laporte, 1985). When measuring physical activity in epidemiology, it is important to ensure that the information being gathered is relevant to the study and that it is accurate. To ensure that information relevant to the study being executed is gathered, the endpoint of the study, the study population, and the outcomes of the study instrument should be considered (Pols, 1998). In order to ensure that a study instrument is accurate, the magnitude of error (both random and systematic within-person error) should be considered by measuring reproducibility (reliability) and validity.

2.1.3 Questionnaires

Questionnaires are the method of choice for assessing physical activity in the majority of epidemiological studies (Pols, 1998). The reason for this preference is that they are practical (they are relatively inexpensive, can be self-administered, and are time efficient, because data on a large timeframe i.e. a week, month or year can be gathered in a single visit), they are applicable (can be designed in an appropriate manner for specific studies) and they are nonreactive (they do not change the behaviour of the study population during data collection) (Kriska, 2005). Questionnaires can differ with regard to length, complexity, type of activities, time period asked about, and method of execution (self-administered or by interviewer; the latter either in person or by telephone). Ideally, questionnaires should address the type, intensity, frequency, and duration of physical activity (Pols, 1998).

2.1.4 Physical Activity and Cancer

Over 180 observational epidemiologic studies have examined the relationship between physical activity and cancer (American Institute for Cancer Research, 2005). The evidence for a beneficial effect of physical activity on cancer risk has been classified as "convincing" for colon and breast cancer, "probable" for endometrial cancer and "possible" for prostate and lung cancer. In a recent review it was found that of the 56 studies conducted on colon and colorectal cancer, 48 demonstrated a reduction in cancer risk among the most physically active participants (American Institute for Cancer Research, 2005). The decreases in risk are substantial, with an average risk reduction ranging from 10-50% (American Institute for Cancer Research, 2005).

Due to the strength of the observed relationship physical activity and colorectal cancer, this cancer is of particular interest in this study. A meta-analysis of the association between physical activity and colorectal cancer using data from 19 cohort studies showed a statistically significant reduction in the risk of colon cancer with recreational activities participated in for both males (RR of 0.79) and females (RR of 0.71) (Samad, 2004). In the same analysis, 28 case-control studies were identified and showed significantly reduced risks of colon cancer in both sexes irrespective of the type of activity (Samad, 2004). Several plausible biological mechanisms for the association between physical activity and cancer risk, specifically colorectal cancer, have been proposed (Friedenrich, 2002) and it is likely that multiple, interrelated actions are involved (Friedenrich, 2002). Although many studies have been done that demonstrate an association between increased physical activity and a decreased risk of colorectal cancer, many questions about this association remain including dose, frequency, timing of physical activity, sex effects and underlying mechanisms (Quadrilatero, 2003).

2.2 Use of Biomarkers/Intermediate Endpoints

Biomarkers of effect and altered function are defined as "processes that are intermediate between exposure and disease" (Perera, 1982) or as "early biological or biochemical changes in the target tissue that result from the action of the carcinogen and are thought to be either a step in the carcinogenic process or correlate closely with that process" (Rothman, 1995). Using biomarkers that lie on the causal pathway for a disease to substitute for the end-stage disease has two major advantages. First, the use of biomarkers makes otherwise rare outcomes more common and second, they make relationships easier to detect. The reason intermediate cancer endpoints are more

common is that while they indicate an increased risk of cancer, they do not invariably lead to the disease. This means that in order to truly understand the nature of the relationship between an exposure and cancer outcome using an intermediate endpoint, the risk of cancer in an individual with the intermediate endpoint needs to be quantified (the percent of individuals who have the intermediate endpoint who get cancer). The reason an exposure/outcome relationship may be easier to detect using an intermediate endpoint is due to the shortened temporal timeframe. By closing the gap between exposure and outcome many factors that may affect the relationship along the way from exposure to outcome are controlled for. Cancer has a long latency and for this reason epidemiological studies using cancer outcomes often require very large sample sizes and long follow-up periods or else these studies ask participants about factors in their life that happened years/decades ago and are subject to recall bias. Utilizing an intermediate cancer end point makes cross-sectional studies more feasible and informative.

Although there is consistent epidemiologic evidence showing a protective effect of physical activity for some cancers, little is known about the mechanisms through which activity exerts its effects (Rundle, 2005). It has been stated that studies utilizing biomarkers of early health effects will help answer remaining questions related to the type and extent of activity needed to elicit a protective effect, whether there are critical time periods in which activity is important, and whether the effect of activity is modified by other factors. This type of information would be very important for designing effective intervention trials and creating population-based prevention programs (Rundle, 2005). This thesis made use of homocysteine as an intermediate end point for cancer and is considered as a possible mechanism for the protective effects of physical activity.

2.3 Homocysteine

Homocysteine ($HSCH_2CH_2CH(NH_2)COOH$ see Figure 2.1 for chemical structure) is a sulphur-containing nonproteinogenic amino acid produced when the essential amino acid methionine is metabolized (Medina, 2001). It exists at a critical biochemical junction in the methionine cycle between S-adenosylmethionine and vitamins B12 and folic acid (Medina, 2001) (see appendix A for model). Although desirable plasma homocysteine concentrations (concentrations that are innocuous and not related to harmful clinical outcomes for the patient) are not clearly defined because the contribution of elevated total plasma homocysteine concentrations to disease is still controversial (Jacobsen, 2001), it is recommended that total plasma homocysteine concentration be maintained below 10 µmol/L (Jacobsen, 1998). Most definitions of 'normal blood levels of homocysteine' are in the range of 5–15.0 µmol/L; 15-25 µmol/L is considered mild hyperhomocysteinemia; 25-50 µmol/L indicates moderate hyperhomocysteinemia; and greater than 50 µmol/L indicates severe hyperhomocysteinemia (Jacobsen, 1998).

The term total plasma homocysteine (tHcy) refers to all forms of homocysteine in the plasma. Only 1-2% of this total occurs as the thiol homocysteine, the rest is in the form of disulfides. Approximately 70% of the homocysteine in plasma or serum is bound to protein through disulfide bonds with protein cysteines, mainly in albumin, and the remaining 25-30% occurs in non-protein bound forms: homocysteine, homocysteine-cysteine disulfide, and more minor amounts of other mixed disulfides (Mudd, 2000).

$$\overset{\displaystyle H}{\underset{\displaystyle CH_2 - CH_2 - SH}{^+H_3N - C - COO^-}}$$
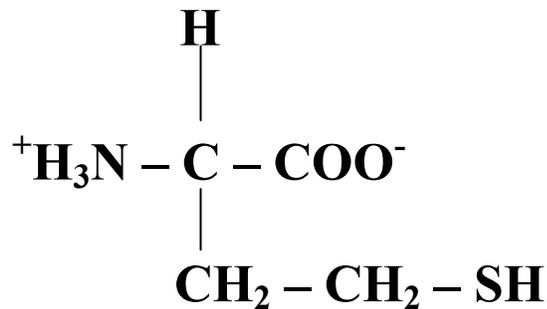
Figure 2.1 Chemical Structure of Homocysteine (Miller 2003)

2.3.1 Homocysteine and Cancer

Elevated tHcy concentration is a suspected risk factor for cancer, cardiovascular disease and several other health effects, including adverse pregnancy outcomes and cognitive impairment in the elderly (Refsum 1998, Medina 2001, Miller 2003). The research in this thesis is more relevant for its potential relationship with cancer than with these other outcomes. Many tumours have a long latency period after initiation and for this reason the relevant window to examine initiators is many years before the age of usual diagnosis. The present study will consider healthy adults between the ages 20 and 50. This is in contrast to cardiovascular or adverse pregnancy outcomes, where contemporary homocysteine concentrations are relevant to the clinical outcomes.

Although studies have observed an association between colorectal, cervical, breast, ovarian, and pancreatic cancer and leukemia and elevated plasma homocysteine levels (Wu 2002, Kato 1999, Corona 1997, Weinstein 2001) only one study exists that demonstrates temporality for the relationship between tHcy and cancer and the cancer type is colorectal (Kato, 1999). Investigations have demonstrated a link between elevated homocysteine status and rate of chromosomal damage in the absence of folate and vitamin B12 deficiencies (Fenech, 1997, 1998). This chromosome damage could

contribute to the increased risk of cancer. The mechanism of DNA damage induced by homocysteine remains to be clarified (Oikawa, 2003); however, disruption in the methionine-homocysteine biosynthesis indicated by elevated tHcy is hypothesized to play a role. A disruption in this cycle is postulated to lead to two potential cancer-causing processes: abnormal DNA methylation, and the production of reactive oxygen species that can lead to oxidative stress (Wu, 2002). See Figure 2.2 for schematic of the methionine-homocysteine cycle.
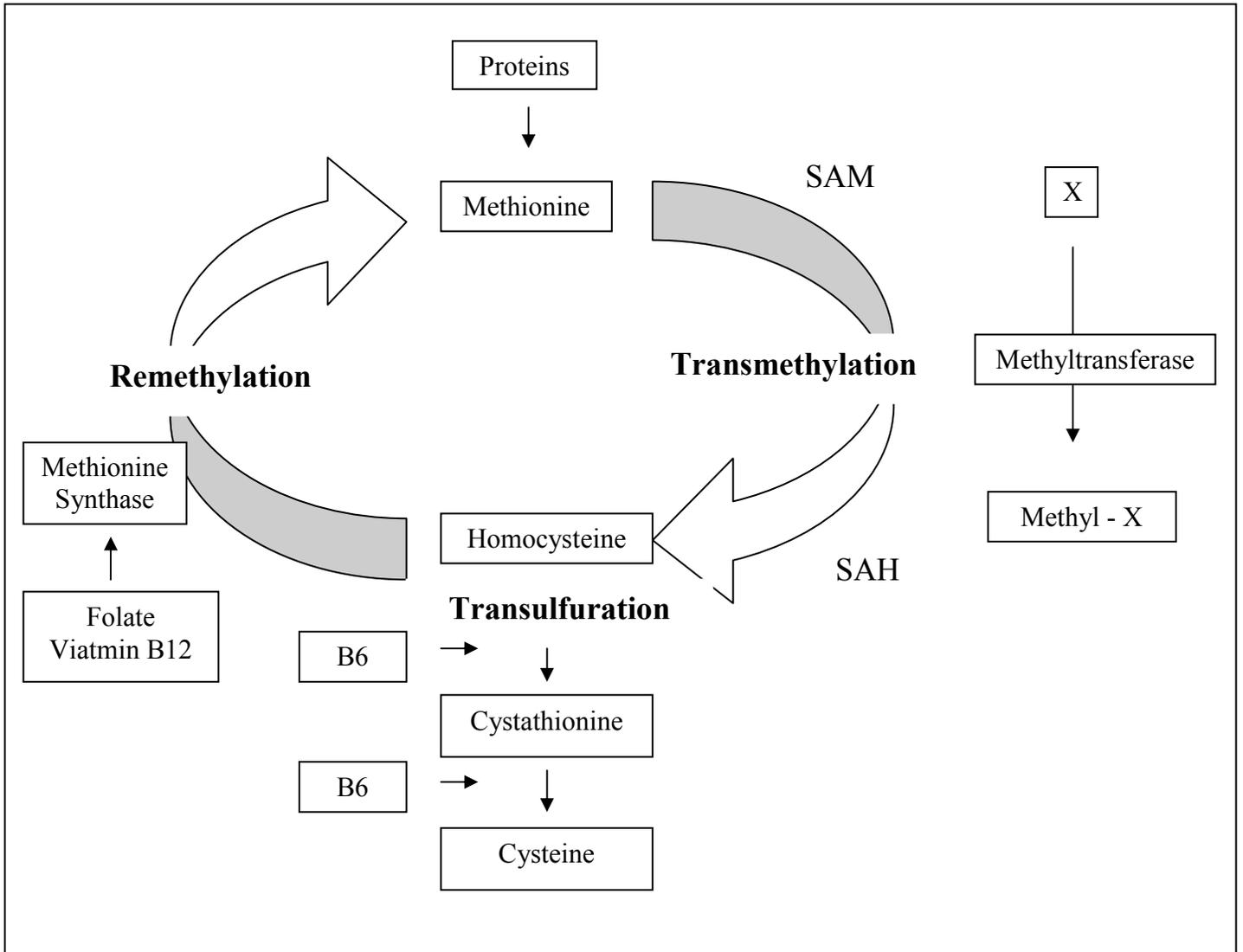
Figure 2.2. Methionine-Homocysteine Biosynthesis (adjusted from Hague 2003) (See appendix A for more detailed figure)

```
                    ┌─────────────────────┐
                    │    Disruption that   │
                    │ inhibits remethylation│
                    └─────────────────────┘
                              ⬇
                    ┌─────────────────────┐
                    │  ↓ Methionine       │
                    └─────────────────────┘
                              ⬇
                    ┌─────────────────────┐
                    │  ↓ SAM              │
                    └─────────────────────┘
                              ⬇
┌──────────────┐    ┌─────────────────────┐
│  Decreased   │    │ Homocysteine redirected│
│  SAM limits  │    │ to transulfuration (does│
│ cystathionine│    │ not follow remethylation)│
│  synthesis   │    └─────────────────────┘
└──────────────┘              ⬇
                    ┌─────────────────────┐
                    │ Limited cystathionine│
                    │ synthesis which limits│
                    │ transulfuration pathway│
                    └─────────────────────┘
                              ⬇
                    ┌─────────────────────┐
                    │ Build up of homocysteine│
                    │       in cell        │
                    └─────────────────────┘
```

homocsyteine released into plasma and quickly auto-oxidizes to form reactive oxygen species

Build up of SAH (powerful methylation inhibitor)

metal-mediated oxidative DNA damage

Hypomethylation of DNA

Dysfunctional DNA repair (activation of protooconcogenes, suppression of tumor suppressor genes)

Carcinogenisis

Fig 2.3   Pathway within methionine-homocysteine biosynthesis that can potentially lead to carcinogensis

13

Although, to date only a single study demonstrates the tHcy-cancer relationship with temporality (Kato, 1999), there are also a number of studies that exist that support the concept that low folate status increases risk of several cancers (including colorectal cancer). These studies are of interest because a disruption in the methionine-homocysteine biosynthesis is thought to explain the folate-cancer relationship as well, and thus, studies of low folate and studies of high homocysteine are looking at the same biological link to increased cancer risk. Studies on tHcy and folate and cancer risk taken together provide a strong body of evidence that dysfunction in methionine-homocysteine biosynthesis plays a role in carcinogenesis.

Low folate levels inhibit the methionine homocysteine cycle and lead to decreased methylation. The relationship between decreased folate and diminished DNA methylation has been observed directly in two human studies (Jacob 1998, Rampersaud 2000). These studies concluded that the hypomethylation induced in circumstances of low folate is 'probably due primarily to the associated rise in tHcy, which in turn increases tissue concentration of SAH' (Choi, 2002). This conclusion is supported by the findings of Yi and colleagues who demonstrated that a chronic elevation in plasma homocysteine concentrations has an indirect and negative effect on cellular methylation reactions through an associated increase in intracellular SAH concentrations (Yi, 2000). Because SAH inhibits biological methylation reactions, increased homocysteine concentrations may directly affect carcinogenesis by diminishing DNA methylation in critical tissues (Choi, 2002). The disruption of methionine-homocysteine biosynthesis also results in a decreased concentration of SAM (the universal methyl donor), which leads to further hypomethylation (Lamprecht, 2003).

14

Decreased methylation has been observed in cancers of the colon, stomach, uterine, cervix, prostate, thyroid and breast (Choi, 2002). Methylation is the biochemical process in which certain molecules transfer or donate a methyl group (-CH3) to another molecule. The universal methyl donor in the body is SAM and when SAM is depleted and SAH is built up one consequence is genomic hypomethylation. DNA methylation is an epigenetic determinant in gene expression, DNA stability, and mutagenesis (Eichholzer, 2001). A very early finding in carcinogenesis, before mutation and deletion events occur, is a decreased level of genomic methylation. This may be due to DNA repair dysfunctions and genomic instability both of which have been found to be the result of hypomethylation (Wu, 2002). This hypomethylation is of particular importance when it occurs within the promoter region of a gene as this can lead to the activation of protooncogenes (Choi, 2002). Upon activation a protooncogene becomes a tumor-producing agent, an oncogene (Todd, 1999). The activation of protooncogenes is not the only concern however, studies have shown that a methyl-deficient diet in rats produced first hypomethylation and then, hypermethylation was reported (Pogribny, 1999). This is of significant consequence because hypermethylation can lead to the silencing of tumor suppressor genes (Choi, 2002). Thus a disruption in the methionine-homocysteine biosynthesis results in aberrant methylation which can lead to both the activation of protooncogenes and silencing of tumor suppressor genes increasing the risk of developing cancer.

A second, less studied way that tHcy could effect carcinogenesis is through oxidative stress. When homocysteine enters plasma it quickly auto-oxidizes and produces highly reactive oxygen molecules (Aguilar, 2004). The endogenous attack on DNA by

oxygen free radicals generates DNA adducts that can be detected in human cells (Wu, 2002). Oxidation of DNA may cause gene mutation and eventually lead to carcinogenesis (Wu, 2002).

2.3.2 Homocysteine and Colorectal Cancer

Kato and colleagues reported in a nested case-control study that the risk of colorectal cancer was significantly elevated among those individuals with a higher total plasma homocysteine concentration even after controlling for folate concentrations. Study subjects were women who had volunteered to participate in the New York University Women's Health Study enrolled in the study between 1985 and 1991. At enrolment, reproductive and dietary data were collected through self-administered questionnaires and 30ml of nonfasting peripheral venous blood was drawn. They considered 'cases' individuals who had been diagnosed with colorectal cancer prior to 1995. Blood samples were taken between 2.4 months and 9.1 years before subjects were diagnosed with cancer and when the seven cases that were diagnosed within 1 year of their blood sample were excluded, the odds ratio did not change (Kato, 1999). Subjects in the highest quartile of tHcy (>12.1μmol/L) had a 70% increase risk of colorectal cancer compared with those in the lowest quartile ($\leq$ 7.9μmol/L) (p=0.09), the mean difference in tHcy between cases and controls was 0.67μmol/L (Kato 1999).

Martinez and colleagues performed a cross-sectional study on men and women who had undergone removal of colorectal polyps. Colorectal cancer is thought to be the result of a multistep process that involves a precursor lesion – the adenomatous polyp (Martinez, 2006). They found in their study a lower recurrence of colorectal adenomas in

subjects with lower homocysteine concentrations indicating a lower risk for developing colorectal cancer (Martinez, 2004).

2.3.3 Homocysteine and Cardiovascular Disease

A meta-analysis using data from 27 studies concluded that an increase of 5 μmol/L of homocysteine in peripheral blood is associated with the same risk of coronary heart disease as an increase of 0.5mmol/L cholesterol (OR of 1.6 for men and 1.8 for women) (Boushey, 1995). A second meta-analysis using data from 72 cohort studies concluded that there is 'strong evidence that the association between homocysteine and cardiovascular disease is causal' and that lowering homocysteine concentrations by 3 μmol/L from current levels would reduce the risk of ischaemic heart disease by 16%, deep vein thrombosis by 25% and stroke by 24% (Wald, 2002). Many biological mechanisms have been proposed to explain the role of homocysteine in cardiovascular disease and it has been suggested that there 'probably is a multifactorial effect whose main consequences are endothelial and vessel wall damage' (Aguilar, 2004).

2.3.4 Homocysteine and Other Adverse Health Outcomes

A review of homocysteine and pregnancy concluded that increased concentrations of plasma homocysteine are associated with fetal neural tube defects, maternal conditions such as pre-eclampsia, and placental disorders such as abruption and recurrent pregnancy loss (Hague, 2003). In addition, elevated homocysteine levels have been linked with cognitive function, ranging from mild cognitive decline to vascular dementia and Alzheimer's disease (Miller, 2003). Since an association has been observed between so many adverse health outcomes, it would be beneficial from a public health perspective to identify modifiable factors that influence homocysteine levels.

2.3.5 Determinants of Homocysteine/Potential Confounders

Factors that have been identified as the strongest determinants of plasma homocysteine concentration in the general population include sex, age, folate intake, smoking status, and coffee consumption (Refsum, 2006). Other factors that have also been found to be associated with total plasma homocysteine concentration are race, alcohol consumption, body mass index, blood cholesterol, blood pressure, and pregnancy (Schneede 2000, El-Khairy 1999, Walker 1999, Michelete 1995, Anderson 1992, Jacobsen 1998, Mora 2006).

More specifically, age and sex are among the most consistent and strongest determinants of homocysteine concentration in adults (Refsum, 2006). Total plasma homocysteine concentration tends to be higher in men than in women (Dierkes, 2001) and it increases with increasing age (Refsum, 2006). Age and sex are related to physical activity as it has been found that males tend to be more active than females (Capersen, 2000), and younger people tend to be more active than older people (Sallis, 2000).

Total plasma homocysteine tends to be higher in individuals who smoke, drink alcohol and drink coffee. A dose response relation has been observed between homocysteine and daily number of cigarettes smoked (Chrysohoou, 2004), and for alcohol and coffee consumption (Chrysohoou, 2004, Refsum, 2006). Also, B vitamins, specifically B12 and folate, are important cofactors of several enzymes involved in homocysteine metabolism; thus dietary intake of these vitamins greatly affects plasma homocysteine levels (Rousseau, 2005). The Hordaland homocysteine study observed a strong inverse relationship between both plasma folate and vitamin B12 and tHcy (Refsum, 2006). The Hordaland homocysteine study also found a positive association

between tHcy and blood pressure and total cholesterol (Refsum, 2006). These variables

are of interest in this study as they have been found to be related to the outcome (tHcy) in

past studies, and are likely related to the exposure of physical activity. Being physically

active is one indicator that an individual has a healthy lifestyle which often includes a

healthy diet and thus higher vitamin intake, lower alcohol, coffee and smoke intake, and

lower blood pressure and blood cholesterol then someone who does not a have healthy

lifestyle.

2.4 Biological Mechanism for Physical Activity altering Homocysteine Concentrations

Although epidemiologic studies have been performed examining the relationship

between physical activity and homocysteine, no biological mechanism for this

relationship has been identified or investigated. This relationship is biologically

plausible, however, as physical activity causes several biochemical changes that could

influence the metabolic pathway of homocysteine. Perhaps most importantly is the role of

oxidative stress. Strenuous long-duration exercise overwhelms our capacity to detoxify

reactive oxygen species and can lead to oxidative stress (Sen, 1995). Whereas moderate-

intensity endurance exercise training can enhance the physiological antioxidant defences

(levels of glutathione, GSH) (Sen, 1999) and reduce the occurrence of oxidative stress.

As well, activity restriction has been observed to significantly compromise oxidative

stress defences (levels of GSH), leaving tissues more susceptible to oxidative damage

(Sen, 1995). Homocysteine is an intermediate in the pathway of synthesis of cysteine

from methionine (the transsulfuration pathway) and is an essential intermediate in the

transfer of activated methyl groups from tetrahydrofolate to S-adenosylmethionine

(SAM) (the activated methyl cycle) (Medina, 2001). When cellular SAM concentration is

low, the synthesis of methyltetrahydrofolate, necessary for remethylation of

homocysteine to methionine, will proceed uninhibited whereas cystathionine synthesis (on the transsulfuration pathway) will be suppressed, thus resulting in the conservation of homocysteine for methionine synthesis (Medina, 2001). When SAM concentration is high, homocysteine is diverted through the transsulfuration pathway because of inhibition of methyltetrahydrofolate synthesis and stimulation of cystathionine synthesis (Medina 2001).

When exposed to oxidative stress, methionine synthase is inactivated and the remethylation cycle can no longer proceed. This results in depressed synthesis of methionine. This leads to the diversion of homocysteine to the transsulfuration pathway. This pathway is incapable of handling the additional homocysteine as the depressed synthesis of methionine will also lead to a decrease in intracellular SAM concentration needed for cystathionine synthesis (the initial reaction of transsulfuration). Thus, the transsulfuration pathway becomes ineffective because of the increased homocysteine burden and decreased concentration of SAM and as a result, homocysteine accumulates in the cell and subsequently is exported into the blood, causing hyperhomocysteinemia. (Selhub, 1999)

This may explain an increase in homocysteine levels in highly trained athletes or directly after long duration strenuous exercise when exercise induced oxidative stress occurs, or in individuals who are sedentary and thus have poor oxidative stress defences. As moderate-intensity endurance exercise training may result in less oxidative stress, homocysteine levels in individuals who take part in this type of activity may be lower.

2.5 Physical Activity and Homocysteine

For this literature review 16 studies were identified that examined the relationship between physical activity and total plasma homocysteine concentration. These consisted of eight cross-sectional studies that used a self-report questionnaire to quantify physical activity, seven randomized clinical trials that randomly assigned participants to different exercise regimes to examine the effect of different training regimens on homocysteine concentrations, and a single case-control study. Although a relationship between high physical activity levels and lower total plasma homocysteine concentrations was observed in several of these studies (Dankner 2002, Hellgren 2005, Nygard 1995, Mora 2006, Gaume 2004), this association is not consistently found in all of the literature (de Bree 2001, Husemoen 2004, Chrysonoou 2004). Due to these inconsistencies and gaps in the literature, questions remain about the exact nature of this association.

2.5.1 Cross Sectional Studies

Of the eight cross-sectional studies found in the literature, four studies reported an inverse association between physical activity and total plasma homocysteine concentration (Mora 2006, Nygard 1995, Hellgren 2004, Dankner 2002), three found no relationship (de Bree 2001, Rousseau 2005, Chrysohoou 2004), and one reported a positive relationship between intense endurance activity and increased plasma homocysteine concentration (Herrmann, 2003). Most of these studies compared individuals in the highest training category with those in the lowest training/no activity category.

Since cross-sectional studies examine a relationship at a single point in time, their inability to establish the temporal sequence of exposure and outcome reliably is often a

limitation of this study design. When examining physical activity and homocysteine levels, however, temporality is of less concern. If a relationship were to be found, it is most probable that physical activity came before altered homocysteine levels since it is unlikely that individuals' homocysteine levels affected their physical activity status.

Some of these cross-sectional studies had potential limitations related to the measurement of folate as a confounder, the choice of study population, and the measurement of physical activity.

Of the eight cross-sectional studies found in the literature on this topic, two were limited to a single sex (Mora 2006, Rousseau 2005) and three addressed this relationship in specific subgroups of the general population that may have different determinants of homocysteine concentrations (elite athletes, older adults, and individuals with type two diabetes) (Hellgren 2004, Rousseau 2005, Dankner 2004). Any relationship found involving homocysteine in an older population (over 60) may not hold for healthy adults aged 20-60. Individuals with diabetes and those who are elite athletes have many biological differences compared to the general population that could affect their homocysteine levels (e.g. renal dysfunction and abnormally high metabolism). Therefore, results from studies on older populations, individuals with diabetes and elite athletes may not be generalizable to the population at large. Finally, as studies on a single sex are not generalizable to individuals of the opposite sex and health risks associated with elevated homocysteine levels are not limited to either sex it is important to study the relationship between physical activity and homocysteine concentrations in both males and females.

There are a number of potential confounders that should be taken into account when examining the relationship between physical activity and homocysteine in order to

elucidate the true nature of this association. Only two studies performed thus far controlled for folate as a confounder (Rousseau 2005, de Bree 2001) and two studies controlled for B vitamins (Rousseau 2005, Dankner 2004). Folate intake is identified as the most important dietary determinant of plasma total homocysteine concentration (de Bree, 2001) and vitamins B12 and B6 are essential factors in homocysteine metabolism. The intake of folate and B vitamins is feasibly related to physical activity as part of a healthy lifestyle. As such, it is important to get an accurate measurement of these covariates and control for their intake in the statistical analysis. Inadequate measurement of confounders reduces the ability of the study to reveal the true relationships between the variables of interest.

All of the cross-sectional studies used self-report questionnaires to quantify physical activity. However, the quality and depth of these questionnaires and the data collected on physical activity was variable. Only two publications in this group addressed the testing done on the questionnaire to confirm reliable and valid measures of physical activity (Mora 2006, Nygard 1995). Two studies represented physical activity categorically with only a single question asking participants to check off one of four boxes being used to measure this variable (Nygard 1995, Hellgren 2004). Categorization introduces a level of misclassification that is not present in studies that measure physical activity as a continuous variable.

Only two cross-sectional studies involved a large, representative population with results that could be generalized to both male and female healthy adults (Chrysohoou 2004, de Bree 2001) and a third study limited the population age to two groups 40-42 and 65-67 (Nygard, 1995). Of these three studies, two found no association between physical

activity and homocysteine levels (de Bree 2001, Chyrsohoou 2004), and one found an inverse association (Nygard, 1995) where higher physical activity levels were associated with lower plasma homocysteine concentrations.

The limitations and conflicting results of previous epidemiological studies examining this relationship highlight the need for further investigation of this important biologically plausible association.

2.5.2 Randomized Control Trials

Seven randomized control trials have examined the relationship between physical activity and plasma homocysteine concentration by randomly assigning individuals to training groups. Six of these trials assigned participants to different intensities and frequencies of endurance activity and one assigned participants to resistance training programs. Of these seven studies, two found that increased activity lowered homocysteine concentration (Vincent 2003, Randeva 2002), two found no association (Husemoen 2006, Boreham 2006) and three found that increased physical activity increased homocysteine concentration (Herrmann 2003, Duncan 2004, Bailey 2000). An advantage of randomized control trials over cross-sectional studies is that there is no question regarding temporality. As well, because individuals are randomly assigned to different treatment or intervention groups, these groups are assumed to be similar with respect to 'unmeasured baseline characteristics' and thus confounders are controlled for in the study design (Rothman, 1998). In a randomized controlled trial it is considered highly preferable to blind the individual who assigns participants to treatment groups, the participants and the individual who will determine the outcome to the treatment or intervention assignment (Rothman, 1998). In trials examining the relationship between

physical activity and homocysteine levels, blinding the assessor of the outcome is not important since an individual's plasma homocysteine concentration is an objective measure and cannot be biased by an individual's knowledge of the treatment group. Also, it is not possible to blind participants to their treatment group as they are assigned to some program of physical activity. Treatment compliance may be a limitation of these studies since individuals assigned to more intense physical activity may be less likely to comply to the assigned exercise regime than those in a less intense exercise program. This type of non-compliance would skew the results towards the null.

The major limitations of the randomized trials included in this review were the use of small sample sizes and very specific populations. Four of these studies had fewer than 21 participants (Randeva 2002, Bailey 2000, Herrmann 2003, Boreham 2005). A small study population limits the statistical power of the study and thus some relationships may have been missed.

Like the cross-sectional studies, some of the study populations in these trials were limited to a specific sex (Randeva, 2002) or age group (Vincent 2003, Herrmann 2003, Boreham 2005) thus limiting the generalizability of the results.

2.5.4 Summary of Studies of tHcy and Physical Activity

In summary, the 16 epidemiologic studies that have examined the relationship between physical activity and homocysteine concentrations have produced conflicting and inconsistent results. Few of these studies made use of a large population of healthy adults, and other potential limitations exist. The proposed study would address the limitations discussed by using a reliable and valid study instrument to quantify physical activity, gathering appropriate information about potential confounders, and involving

participants who are healthy adults from both sexes. This study will add to the body of evidence on this topic and will help to determine the nature of this relationship.

2.6 Outcome Measurement Issues:

Stability of Total Plasma Homocysteine Over Time (Within person variability)

To ascertain the study outcome measurement (plasma homocysteine concentration), a single fasting blood sample was obtained in one 10mL lavender K2EDTA tube. EDTA tubes must to be inverted 8-10 times to mix the anticoagulant with the blood and prevent clotting. Tubes are then centrifuged in a swinging bucket centrifuge (preferred type of spin for gel separation tubes) for ten minutes at a speed of 1100 to 1300 relative centrifugal force (RCF) according to manufacturers' recommendations. In order to ensure accuracy of tHcy measurement blood samples were placed on ice immediately after being drawn and were centrifuged and the plasma was frozen $\leq$ 30 minutes after collection to prevent the release of free homocysteine from erythrocytes (American Society of Human Genetics).

To address the ability of a single measurement to represent an accurate assessment of the outcome of interest, it is important to consider the within person variability of this variable. Clarke et al. reported in a study in which seven blood samples were taken from a cohort of 96 participants over the course of one year, that an individual's total plasma homocysteine concentrations were relatively constant over the period of one year, with little seasonal variation. The reliability coefficient (R) was 0.88 (Clarke, 1998). Within-individual differences in homocysteine concentration were with a small coefficient of variation (CV ratio of standard deviation to the mean) of 9%, whereas the between-person concentrations differed much more substantially with a CV

= 24% (Clarke 1998). In a second study there was excellent agreement between four homocysteine measurements taken for one individual one week apart over the period of one month. (Garg, 1997). The results from both of these studies suggest that total plasma homocysteine concentrations are stable, and that a single blood sample a sufficiently reliable measurement to represent this outcome.

2.7 Exposure Measurement Issues

Study Instrument

'One of the greatest challenges to scientific inquiry on the association between physical activity and chronic disease is the measurement of physical activity disease exposure' (McTiernan, 1998). In past epidemiological studies, several study instruments have been used to quantify physical activity and, as has been discussed, questionnaires are the method of choice in the majority of epidemiological studies (Pols, 1998). In this study the short-format, self-administered version of the International Physical Activity Questionnaire (IPAQ) (Craig, 2003) was chosen to quantify participant's physical activity for the month previous to their blood collection (see appendix). This questionnaire is made up of seven questions regarding an individual's activity. It was chosen from a large number of available questionnaires because it is concise, it has gone through rigorous international testing for short-term test-retest reliability and validity, it was designed to be used by adults aged 18-65 years, and it was developed by an international group of physical activity experts. As well, this questionnaire has a predefined scoring protocol that will be used for the statistical analysis of the data collected. Using the IPAQ study participant's cumulative physical activity will be represented as MET-minutes per week, which can be examined as a continuous variable

or can be categorized. A metabolic equivalent or MET is the ratio of the work metabolic rate to the resting metabolic rate. One MET is defined as 1 kcal/kg/hour and is roughly equivalent to the energy cost of sitting quietly. The results of a reliability and validity study of the IPAQ (both the long and short form) indicate that the IPAQ is at least as good as other established self-report activity measures (Craig, 2003).

The *a priori* hypothesis of this study was that the relationship between physical activity and tHcy was U-shaped. That is that sedentary people would have higher tHcy concentrations, moderately or 'healthily' active people would have lower tHcy concentrations, and highly/intensely active people would have higher tHcy concentrations again. This U-shaped relationship was observed by a group conducting the 'Hordaland Homocysteine Study' a population based study which included more than 18000 men and women in Norway (Nygard, 1995). As well, this U-shape is plausible from a biological point of view.

Before conducting this study it was recognized that physical activity and physical fitness, although closely associated, are different measures and a choice had to be made as to what the exposure of interest would be. This study measured physical activity rather than physical fitness for two reasons. First the literature points to a stronger relationship with physical activity than fitness. A study by Duncan et al observed no relationship between plasma homocysteine and VO_2 max (the maximum amount of oxygen (in millilitres) one can use in one minute per kilogram of body weight, the gold standard for measuring cardiovascular fitness) at baseline and six months into their study in which they assigned participants to different exercise regimes (Duncan, 2004). Wright et al came to the same conclusion, finding in their study that plasma homocysteine and VO_2

max were not associated in healthy young men following acute exercise (Wright 1998). Secondly, the larger study (from which the data for this thesis is coming) was collecting information on physical activity, and neither the means nor opportunity to measure fitness was available. Furthermore, physical activity is of interest from a public health perspective as it is a modifiable risk factor thus, if a relationship is found with elevated homocysteine levels recommendations can be made to increase participation in activity.

2.8 Summary of Rationale

Colorectal cancer is the 4th leading cause of cancer death worldwide and in 2000 was found by the World Health Organization to account for 7.9% of all cancer deaths (Quadrilatero, 2003). In developed nations colorectal cancer mortality is only second to lung cancer. The geographic distribution of colorectal cancer incidence suggests that lifestyle factors, including physical activity, play a role in its etiology. Ecological support for a lifestyle role in the etiology of colorectal cancer comes from evidence that colorectal cancer risk of immigrants approaches the risk of the host country, regardless of the direction of risk (Quadrilatero, 2003). Evidence from epidemiological and animal studies indicates that physical activity plays a role in the etiology of colon cancer (Quadrilatero, 2003). The biological mechanism that mediates this relationship is yet unknown. This study investigated the functionality of the methionine-homocysteine biosynthesis to partly explain the physical activity colon cancer relationship.

Homocysteine is an amino acid that plays a role in the methionine-homocysteine biosynthesis – dysfunction in which is postulated to lead to several cancer causing mechanisms. In this study homocysteine is used as a biomarker of the functionality of this cycle, with elevated levels of total plasma homocysteine indicating a dysfunction.

Using a biomarker that lies on the causal pathway for a disease as a substitute for a classical endpoint has three major advantages: makes otherwise rare outcomes more common, makes relationships between exposures and outcomes easier to detect, shortens temporal timeframe.

Chapter 3.0: Methods

3.1 Objective

To determine the relationship between physical activity and total plasma homocysteine concentration (tHcy).

3.2 Hypothesis

The hypothesis driving this research was that a U-shaped curve would describe the relationship between physical activity and tHcy.

3.3 Basic Study Design and Methods

This thesis research was nested within a larger cross-sectional study funded by CIHR, which is examining the relationship between environmental and lifestyle factors and biomarkers of methionine-homocysteine biosynthesis. The study investigators are Drs. Will King (PI), Thomas Massey and Ian Casson from Queen's University, Dr. Linda Dodds from Dalhousie, and Dr. Sherry Perkins from the University of Ottawa. The aim of the larger study is to determine the relationship between exposure to disinfection by-products and markers of metabolic processes in methionine-homocysteine biosynthesis, and includes physical actively as a co-variate. This thesis involved 232 healthy volunteers recruited from study centres in Kingston, Ontario and Halifax, Nova Scotia and was based on the first year of the larger study's data. As part of the larger study, subjects provided a blood sample for genomic and biochemical analysis, and completed a questionnaire. The outcome being examined in this study is plasma total homocysteine concentration which is one of the biomarkers being measured in the larger study.

3.4 Subject Recruitment /Study Population

Subject recruitment and data collection/entry for the larger study started in

September 2006. Subjects were recruited through posters and pamphlets placed in buildings around the three study centres. The target population was male and female subjects aged 20 to 50 recruited in approximately equal numbers within 10-year age intervals. Subjects aged 20 to 50 were targeted as this represents a meaningful window in terms of the postulated biologic mechanisms underlying the cancer hypothesis in this research. In addition, in age groups younger and older than this range natural biologic processes are strong determinants of the outcomes of interest such that variation due to the exposures of interest would not be detectable. Subjects who had health conditions that might be related to the outcome measures (e.g. history of angina or other vascular disease, cancer, or diabetes), who were pregnant or had given birth in the previous year were excluded from the study sample. All participants were given a $30 honorarium upon completion of participation.

3.5 Data Collection

Once an individual expressed interest in participating in the study through contacting the study office an appointment was made for the blood draw and to fill out the questionnaire. The study coordinator met with each participant at the time of the appointment and went over the study information pamphlet and consent form. The study information pamphlet explained to participants the hypothesis of the study (in lay language), the background and relevance of the study to society, and the tasks and procedures involved in participation. Informed choice was employed, with one copy of the consent form retained by the subject. The consent form informed the subjects of the procedures involved in participation in the study including blood draw, blood pressure assessment, questionnaire, and home water sample. All participants signed an informed

consent document prior to the blood draw and filling out a questionnaire. Once consent was attained and the participant had been sitting quietly for five minutes their blood pressure was measured with an automated non invasive blood pressure monitor which measured blood pressure six consecutive times with a one minute break between each measurement. The initial measurement was dropped and the average of the remaining five measurements was recorded as blood pressure. After blood pressure measurement was taken, the study coordinator continued with the venipuncture to collect the blood sample. Once the blood sample was obtained participants filled out the questionnaire.

3.5.1 Physical Activity Measurement

The International Physical Activity Questionnaire (IPAQ) was used to quantify participant's physical activity (see appendix E). The questionnaire contained seven questions asking participants the number of days per week and hour/minutes per day that they took part in vigorous activity, moderate activity, walking and sitting in a usual week in the past month. All answers were converted from hours and minutes to just minutes and, as was recommended by the IPAQ scoring protocol, all walking, moderate and vigorous time variables exceeding 4 hours or 240 minutes were re-coded to be equal to 240 minutes. This rule permits a maximum of 28 hours of activity in a week to be reported for each category of physical activity. The data from the questions were used to assign each participant a physical activity score in units of Metabolic Equivalent (MET)-minutes-per week, which are calculated as explained in the following paragraph. The goal of this scoring was not to assign each participant their actual MET-minutes-per week, but to rank participants by participation in physical activity.

The creators of the IPAQ provide a scoring protocol to assign participants MET

min/wk using their answers to the questionnaire. An average MET score was derived for each type of activity by the IPAQ creators using the Ainsworth et al. Compendium of Physical Activity (Ainsworth et al 1993). The following values were used for the analysis of IPAQ data: Walking = 3.3 METs, Moderate Physical Activity = 4.0 METs and Vigorous physical activity = 8.0 METs. The total physical activity MET-minutes/week was calculated using the following formula:

Total physical activity MET-minutes/week = 3.3*walking minutes*walking days + 4.0*moderate-intensity activity minutes*moderate days + 8.0*vigorous-intensity minutes*vigorous-intensity days.

To examine this variable categorically the IPAQ scoring protocol was used once again. It is suggested in the protocol, that physical activity be divided into three categories (Inactive, Minimally Active and Health Enhancing Physically Active (HEPA)) which take into account both total volume and the number of days of physical activity. The criteria for these categories can be found in appendix D. To investigate physical activity fully, two other forms of categorization was used: by quintiles of total MET score and dichotomously by vigorous activity.

3.5.2 Venipuncture and Blood Analysis

To ensure that recent food intake did not affect the levels of homocysteine being measured, the phlebotomy procedure took place between the hours of 8 and 10 am after a twelve-hour overnight fast (including the avoidance of alcohol and coffee consumption) (Rousseau 2004, Chrysohoou, 2004). In addition, blood samples were taken at least 12 hours after the last exercise session (i.e. the day before the blood sampling subjects should only be exposed to light training) (Rousseau, 2004) as it has been found that

intense endurance exercise causes an immediate, temporary increase in homocysteine levels that lasts up to 24 hours after the end of exercise (Herrmann, 2003). Blood samples were drawn with suitable vacutainers (SST and EDTA tubes) from the median antecubital vein. Samples were immediately put on ice and subsequently centrifuged at 3300 rpm for 10min at room temperature. Serum and plasma were then separated into 5ml aliquot tubes and then stored at – 80 degrees Celsius until assayed. Blood samples were shipped to a the Department of Pathology and Laboratory Medicine of the Ottawa General Hospital and were analyzed under the supervision of Dr. Sherry Perkins.

3.5.3 Potential Confounders

Information regarding potential confounders was collected through the study questionnaire, measured directly at the time of the blood draw, or was determined through analysis of the blood drawn or water sample. Potential confounders included were factors that had been identified in previous literature as being determinants of homocysteine. Continuous variables were looked at in relation to tHcy both continuously and categorically (by quartiles) in order to determine the best representation of these variables for the final model.

Blood pressure was measured directly at the time of the clinic visit. Weight and height were recorded to calculate body-mass index (BMI), which is the ratio of weight to the square of the height (kg/m^2). Age, sex, smoking, alcohol, race, and coffee consumption were assessed through the questionnaire.

Participants were placed in one of three categories for smoking status: 'never smoked', 'past smoker' (at least 1 cigarette a day for six months or more), or 'current smoker' (at least one cigarette a day for the past month). This method of categorizing

smokers is similar to the categories used in the 'ATTICA' study of associations between smoking, physical activity, dietary habits and plasma homocysteine levels (Chrysohoou, 2004). Although a dose-response relationship has been observed between number of cigarettes smoked/day and tHcy in past studies (Refsum 2006, Chrysohoou, 2004) this was not controlled for in this study due to the low number of current smokers (n=19) that took part.

Data on the frequency of drinking alcohol in the past month and the number of alcoholic beverages consumed during each of those instances was gathered through the questionnaire for each participant. This information was used to calculate the average number of drinks per week for the past month.

In order to classify participants according to race, each individual was asked to check one of White, Chinese, South Asian, Black, Native/Aboriginal, Arab/West Asian, Filipino, South East Asian, Latin American, Japanese, Korean or other, for both their mother and their father. Participants who indicated that both their mother and father were White were classified as 'White', participants who indicated that both their mother and father were one of Chinese, South Asian, South East Asian, Japanese, or Korean were classified as 'Asian', participants who listed both parents as Black were classified as 'Black', and all other participants were classified as 'other'.

Information was gathered about both coffee containing caffeine and decaffeinated coffee consumption at home and at work. This data was combined to create a cups/week continuous variable for coffee consumption.

Levels of lipids, triglycerides, and vitamin intake (B12 and folate) were measured in blood samples.

Disinfection by-products are formed during water treatment when chlorine reacts with naturally occurring organic matter in the source water. It has been hypothesized that exposure to these chemicals interferes with the methionine-homocysteine biosynthesis. For the purpose of this study a measure of residential trihalomethanes (THMs) (the primary disinfection by-product) is used as a proxy to represent all disinfection by-product exposure. This study had access to accurate data on residential THMs, as this was the exposure measurement of the larger study. This was measured through analysis of water samples and was included as potential confounder.

3.6 Available Sample Size

At the start of the study a projected sample size of 250 was based on the estimated number of subjects that would be recruited for the main study in a time frame that would allow data analysis for this thesis to begin in the spring of 2007.

3.7 Detectable effects

In the analysis of the linear effects of standardized MET score on standardized total plasma homocysteine concentration for a sample size of 250, with significance of 0.05 and power of 80%, the detectable effect is dependent on the anticipated correlation between standardized MET and standardized total plasma homocysteine concentration (Dupont and Plummer 1998).

The detectable effects for MET-homocysteine correlations of 0.01 and 0.4 with the parameters listed are slopes of 0.18, and 0.15 respectively. Relationships identified between homocysteine levels and demographic and lifestyle factors support that differences of this magnitude are a reasonable target. Differences of over 50% of a standard deviation in total plasma homocysteine concentration have been observed for

contrasts of age, sex, smoking, coffee consumption, body mass index and blood

cholesterol and levels of homocysteine.

3.8 Design Effect

The approach of recruiting subjects from selected institutions serving a small

number of municipalities is analogous to cluster randomization, in that observations

within a cluster (institutions) may be more alike than observations selected entirely at

random. As a result, observations may lack independence and variance estimates from

standard statistical analysis are under-estimated (Donner, 2000). One factor affecting

non-independence is the correlation of exposures within clusters and this is not

anticipated to be strong for physical activity as an exposure. Therefore, an inflation factor

was not used in the sample size calculation. Nevertheless, a design effect will be

considered in the analysis via the inclusion of a random effects parameter representing

centre (or institution) in a mixed regression model – and it is recognized that the power of

the study in the calculations above may be slightly over-estimated if a meaningful design

effect exists. For example, if a variance inflation factor of 1.2 was observed in collected

data, the effective sample size would be 208 and the detectable effects would be 0.20 and

0.17 for correlation of 0.01 and 0.4 respectively.

3.9 Data analysis

3.9.1 Descriptive Statistics

All data analysis was done using the SAS statistical program. Before investigating

the relationship between the exposure and outcome of interest all variables identified *a*

priori to be potentially important in elucidating this relationship were considered. The

distribution (mean, median, mode, and standard deviation) of all continuous variables

were determined and frequencies of categories were found for categorical variables. All

continuous variables were considered both continuously and categorically (by quartiles or

clinical cut points) in order to identify the method that best predicts the outcome and thus

best controls for the effect of that variable. To determine which representation of these

variables would better control for confounding they were regressed on tHcy both

continuously and categorically and the p-values were taken from these univariate models

and compared. The p-values were used to determine which representation of each

variable had the strongest association with tHcy, and this representation was then used for

the backwards-stepwise assessment that followed. For categorically represented

variables, any categories that contained less than 5% of participants (10 people) where

combined.

Physical activity was examined continuously as well as categorically. Three

different ways of dividing the physical activity scores into categories were used 1) the

IPAQ-designated categorises described in appendix, 2) quintiles of MET-min/wk and 3)

dichotomously – individuals who participated in any vigorous activity versus those who

did not. The IPAQ-designated categories were used as they take into account both type of

activity done (moderate, vigorous or walking) and the number of MET-min/wk, however

this method groups all those who have performed what has been labelled 'Health

Enhancing Physical Activity' together. In this study the participants who participated in

the highest levels of (or the most) physical activity were postulated to be different then

those who perform a 'healthy' amount of physical activity because of the U-shaped curve

observed in past studies (Nygard, 2001). Thus the second categorization by quintiles of

MET-min/wk was used. An effect of vigorous activity on homocysteine levels has been

observed in past studies and for this reason a separate examination of this type of activity was of interest.

As our measurement of tHcy provided a continuous variable and to limit misclassification the outcome (tHcy) was examined as a continuous variable.

3.9.2 Relationships between covariates

Pearson product moment correlation coefficients were used to compare continuous variables, an F-statistic and p-value were used to examine relationships between dichotomous/categorical variables with continuous variables. Comparisons between categorical variables were performed with the chi square test. The relationships between covariates were of interest, as strong relationships between covariates will have an affect on model building. For example if two variables both showed a strong univariate relationship with tHcy and one lost its significance when included in a multivariate model, this change in significance could be explained by looking at covariate relationships and unveiling any collinearity present. As well, many relationships between the covariates included in this study have been observed in previous studies. Confirming that these relationships exist in the data being used for this study provides confidence that the data collected follows known trends.

The relationship between the covariates and the main exposure (physical activity) and outcome (tHcy) were considered in a similar manner to the covariate relationships (Pearson correlation and F-statistics). The crude relationship between tHcy and physical activity was examined by looking at the Pearson Correlation Coefficient, looking at the means of tHcy for different categories and using the p-value from the F statistic to determine if differences were significant and, creating a simple linear regression model

regressing each representation of physical activity individually on tHcy while controlling for age and sex.

3.9.3 Model Building

The basic analysis used for this project was a least squares regression model considering a linear relationship between physical activity MET score and tHcy level (e.g. tHcy = a + b(METs)). A quadratic term for physical activity was included in the model and was tested for significance in order to explore the possibility of a non-linear relationship between physical activity and tHcy.

In order to build a model to quantify the relationship between the physical activity variable and homocysteine while controlling for potential confounders a multiple linear regression model was formed. To create the most parsimonious model to describe this relationship the backward elimination procedure was used. All covariates were included in a model to predict tHcy. The partial F-statistic and accompanying p-value for this statistic was observed for each variable. The variable with the highest p-value was removed. This process was repeated until all variables had a p-value of <=0.20. Backward elimination (BE) is based on the idea that if a variable does not affect the outcome of interest it cannot be a confounder. This process can lead to the under-selection of important confounders if the cut off p-value is the conventional 5%, thus a liberal p-value (20%) was used in this study (Budtz-Jorgensen, 2006). BE as method of confounder selection was used over the change in estimate method in order to provide a consistent model for all of the representations of physical activity. The change in estimate method is often recommended over p value-based methods of confounder selection because it examines the change in the exposure effect estimate (thus considering the

effect of a variable on the relationship of interest – a better definition of a confounder) (Budtz-Jorgensen, 2006). To provide confidence in the confounder selection method used, the change in estimate method (Rothman and Greenland, 1998) was done in a sub-analysis for one representation of physical activity and the point estimates of effect and confidence intervals for the physical activity variable attained by BE and change in estimate method were compared.

3.10 Ethical Considerations

All participants in this study were volunteers who were fully informed of the study hypothesis. All identifying information was held in strict confidence in locked filing cabinets. Only identification numbers were used when biological specimens were transported. Computerized data files contained no identifying information; it is not possible to identify individuals in any report. The larger study was granted ethics approval at Queens University.

Chapter 4.0: Results

The data analysis for this study was done using 'Statistical Analysis System' (SAS). In this section descriptive statistics are provided for all variables considered in this research, followed by bivariate relationships between potential confounders, potential confounders and exposure of interest (physical activity), and potential confounders and the outcome (tHcy). The crude relationship between physical activity and tHcy is then presented, followed by confounder selection and the full model. In the final sub-section, sensitivity analyses are presented to examine the confounder selection method used and the representation of the outcome of interest (tHcy).

4.1 Descriptive Analysis

4.1.1 Descriptive Statistics for Categorical Variables

Table 4.1 presents data that summarizes the frequencies and percentages of all categorical covariates. The total sample size of this study was two hundred thirty, made up of seventy-seven (33%) males and one hundred fifty-five (67%) females. As the results indicate the study sample was composed of primarily Caucasians, who were non-smokers with a healthy BMI. As the Kingston centre started data collection first, three-quarters of the study sample for this research are from Kingston.

4.1.2 Descriptive Statistics for Continuous Variables

Table 4.2 presents the descriptive statistics for all continuous covariates. Basic summary statistics are presented for each variable including mean, median, stand deviation and 25th and 75th percentiles. Only 4% (9 individuals) of the study population had what is considered low folate (<15 nmol/L), 1.7% (4 individuals) had what is considered clinically to be high blood pressure ($\geq$ 140 mm Hg systolic pressure, or $\geq$ 90

mm Hg diastolic pressure). The age distribution was relatively even within 10 year age

intervals.

Table 4.1   Descriptive Statistics for Categorical Covariates

Categorical Variables	%	# (n=230)
Sex		
Male	32.61	75
Female	67.39	155
Smoking		
Never	73.04	168
Past	15.22	35
Current	11.74	27
Race		
white	84.35	194
Asian	9.13	21
other	6.52	15
Second-hand smoke exposure		
<1	81.30	187
1-4 hours per week	13.04	30
>5 hours per week	5.65	13
BMI		
<18.5	2.17	5
18.5-25	63.04	145
25-30	24.78	57
>30	10.00	23
Center		
Kingston	76.52	176
Halifax	23.48	54

Table 4.2 Descriptive Statistics for Continuous Variables

Variable	Mean	Standard Deviation	Median	25th	75th
Folate (nmol/L)	31.86	25.48	28.80	22.25	36.10
Residential THM (µg/L)	22.74	18.35	16.20	10.30	36.00
BMI (kg/m^2)	24.47	4.43	23.54	21.62	26.30
Alcohol Consumptions (drinks/wk)	3.89	5.60	2.00	0.25	5.00
Systolic Blood Pressure (mmHg)	105.63	9.80	105.00	99.00	112.00
Diastolic Blood Pressure (mmHg)	68.81	8.40	68.81	63.00	74.50
Age (years)	32.17	9.28	30.00	24.00	41.00
Triglycerides (mmol/L)	0.88	0.50	0.72	0.55	1.05
Vitamin B12 (picomol/L)	239.84	95.23	232.10	167.90	292.95
Coffee Consumption (cups/wk)	9.17	11.30	6.00	0	14.00
HDL Cholesterol (mm/L)	1.33	0.33	1.29	1.11	1.48
LDL Cholesterol (mm/L)	2.96	0.74	2.91	2.43	3.40
Cholesterol (mm/L)	4.45	0.81	4.40	3.90	5.00

4.1.3 Physical Activity

The creators of the IPAQ suggest truncating data when walking, moderate or vigorous time variables exceed 3 hours or 180 minutes per day. This rule permits a maximum of 21 hours of activity in a week to be reported for each category. After applying this truncation to the data the physical activity scores ranged from 0 MET-min/wk to 12078 MET-min/wk. The median and mean scores were 2319 MET-min/wk and 2954 MET-min/wk respectively and the overall distribution of scores demonstrated a slight negative skew. In table 4.3 the distribution of the physical activity scores are given in all representations: categorically using the IPAQ designated categories, quintiles, and dichotomously based on participation in vigorous activity are presented. The continuous distribution of physical activity scores demonstrates a positive skew. Figure 4.1 provides a histogram of the continuous physical activity scores with the normal distribution overlaid and basic summary statistics.

Fig 4.1   **Distribution of physical activity scores**
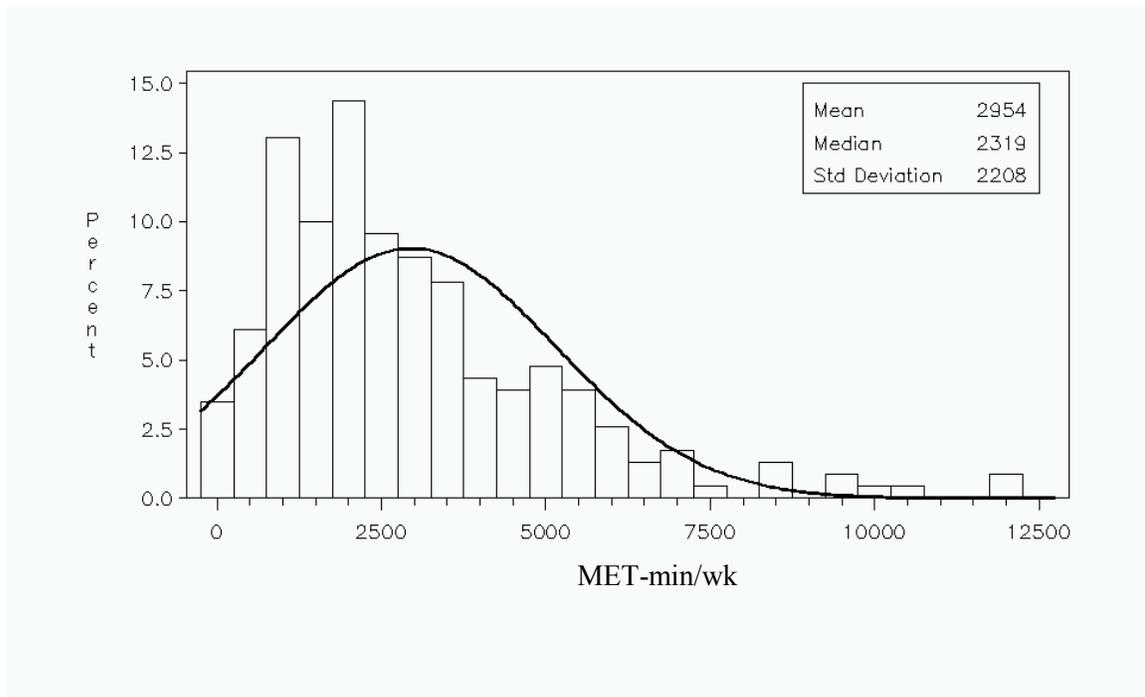
Table 4.3   Univariate Analysis of Physical Activity

Category	% Participants (n=230)
IPAQ	
Inactive (median= 396)	8.26 (19)
Minimally Active (median=1758)	49.57 (114)
HEPA Active (median=4386)	42.17 (97)
QUINTILES	
(<1140 MET min/wk)	19.57 (45)
(1140-2013 MET min/wk)	20.43 (47)
(2013-2862 MET min/wk)	20.87 (48)
(2862–4452 MET min/wk)	19.13 (44)
(> 4452)	20.00 (46)
VIGOROUS	
Vigorous	73.04 (168)
No Vigorous	26.96 (62)

4.1.4 Total Plasma Homocysteine Concentration Distribution

The tHcy concentrations found in this study ranged from 3.91 µmol/L to 17.53 µmol/L with a mean of 8.02 µmol/L. A histogram of the distribution of tHcy stratified by sex, along with descriptive statistics can be found in figures 4.2 and 4.3. The distribution of tHcy concentrations was checked for normality using the Shapiro-Wilk test. This test provided a W of 0.94 and p <0.01 indicating that the data was not normally distributed.
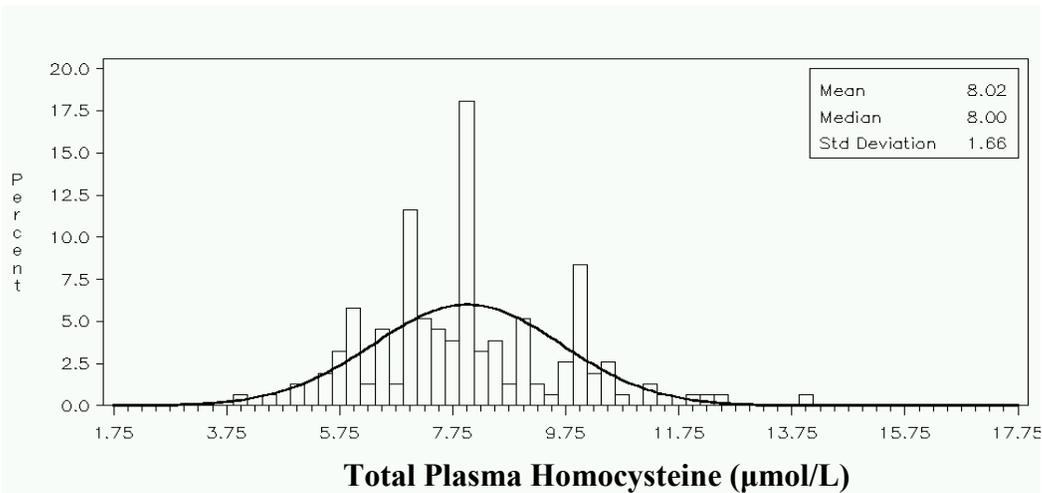
Fig 4.2 Female Total Plasma Homocysteine Distribution
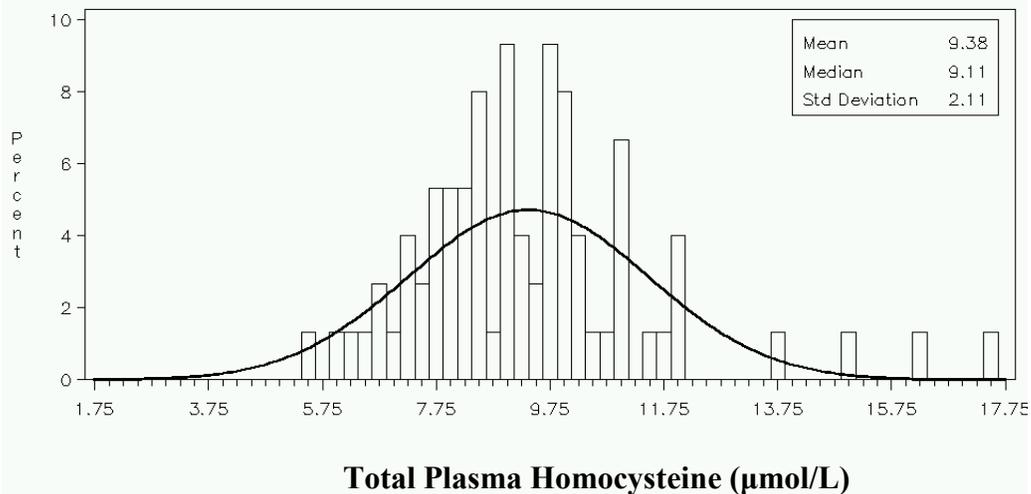
Fig 4.3 Male Total Plasma Homocysteine Distribution

4.2 Bivariate Analysis

The relationships between covariates were examined to provide an understanding of what collineararities exist, which in turn helps to understand why variables may not be included in the final model. As well, many relationships have been observed in previous studies between two or more of these variables and it is of interest to see if these relationships hold within this data set.

4.2.1 Covariate Correlations and Interrelationships

Table 4.4 presents the p-values from the appropriate tests to look at the relationships between covariates. For the relationship between continuous and categorical variables, the p-values from the F statistics from simple linear regressions are provided. Chi-square tests were used to examine associations between categorical variables and Pearson's Product correlation coefficients were found to examine the relationships between continuous variables. Statistically significant relationships were found between many of the covariates; these p values are highlighted in the table.

4.2.2 Physical Activity and all other Variables

In order to examine the relationship between physical activity and the other variables being considered in this study the continuous representation of physical activity was used (MET-min/wk). The relationship between physical activity and the continuous variables was examined using Pearson's product-moment correlations (see table 4.5). A statistically significant relationship was only observed between physical activity and coffee consumption, with a Pearson Correlation Coefficient of 0.15 or r^2 of 0.0225. This means that 2.2% of the variation of physical activity can be explained by variation in participant's coffee consumption. Alcohol consumption had a weak positive correlation with physical activity.

Table 4.4 Relationships between Covariates:

	BMI	Age	Cholest	Tri	HDL	B12	Folate	Sys	Dia	LDL	THM	Drink	Coffee	Sex	Smoke	Race	2nd smk	Center
BMI		<0.01	0.04	<0.01	<0.01	0.99	0.67	<0.01	<0.01	<0.01	0.04	0.99	0.01	0.13	0.02	<0.01	<0.01	0.20
Age			<0.01	0.26	0.37	0.78	0.66	<0.01	<0.01	<0.01	0.03	0.03	<0.01	0.03	0.04	0.20	0.53	<0.01
Cholesterol				<0.01	<0.01	0.78	0.66	0.23	0.02	<0.01	0.20	0.57	0.01	0.07	0.24	0.10	0.34	0.19
Triglycerides					<0.01	0.99	0.70	<0.01	<0.01	<0.01	0.18	0.04	0.83	0.49	0.03	0.70	0.57	0.93
HDL						0.53	0.50	<0.01	0.27	0.15	0.17	0.86	0.75	<0.01	0.99	0.49	0.19	0.16
B12							0.10	0.86	0.42	0.31	0.04	0.20	0.77	0.01	0.68	0.28	0.99	0.14
Folate								0.63	0.40	0.50	0.28	0.55	0.30	0.16	0.27	0.76	0.24	0.09
Systolic BP									<0.01	0.04	0.77	<0.01	<0.01	<0.01	0.19	0.03	0.01	0.99
Diastolic BP										0.01	0.22	0.03	<0.01	0.04	0.14	0.05	0.05	0.09
LDL											0.06	0.70	<001	0.60	0.36	0.04	0.56	0.04
THM												0.32	0.17	0.04	0.13	0.08	0.06	<0.01
Alcohol													0.55	<0.01	0.09	0.28	0.13	0.45
Coffee														<0.01	<0.01	0.95	<0.01	0.40
Sex															0.13	0.56	0.08	<0.01
Smoking																0.02	<0.01	0.29
Race																	0.42	0.01
2nd-hand smoke																		<0.01
Center																		

* p value from Pearson Product Moment correlations for pairs of continuous variables, from F statistic from a linear regression for continuous and categorical variables, and from chi square test for relationships between two categorical variables

Table 4.5   Relationships between Physical Activity and Continuous Covariates

Continuous Variables	Pearson Correlation Coefficient	p value
BMI	-0.04	0.54
Age	-0.08	0.24
Cholesterol	0.01	0.86
Triglycerides	-0.02	0.78
Coffee	0.15	0.03
B12	0.03	0.67
Folate	-0.04	0.59
Systolic BP	-0.02	0.74
Diastolic BP	-0.09	0.19
Alcohol Consumption	0.13	0.06
Chlorination by-products	0.02	0.76
HDL Cholesterol	-0.03	0.70
LDL Cholesterol	0.02	0.73

An F-statistic from linear regressions (regressing the categorical variables on physical activity total MET score) were used to determine if the relationship between physical activity and a categorical variables was statistically significant. The p-values for the F-statistic and the mean MET-min/wk score are presented in table 4.6. A statistically significant relationship was observed between physical activity and sex. A weak relationship was observed between physical activity and smoking status. Some of the relationships observed are in the opposite direction to what might be expected with females being more active than males and those who consumed the highest amounts of coffee and alcohol and being current smokers being the most active. These associations can likely be attributed to the age of participants (with the younger people being more active and also drinking both alcohol and coffee and smoking more than older people).

Table 4.6 Relationships between Physical Activity and Categorical Variables

Variable	Mean MET-min/wk (std dev)	P value*
Sex		0.02
Male	2724.4 (2196.4)	
Female	3428.4 (2183.98)	
Smoking		0.08
Never	3027.9 (2168.5)	
Past	2235.7 (1866.6)	
Current	3425.4 (2688.0)	
Alcohol Consumption – drinks per week		0.79
(0- $\leq$ 0.25)	2709.3 (2390.0)	
(> 0.25 - $\leq$ 1.875)	2923.1 (2282.2)	
(>1.875-$\leq$ 4.5)	2996.2 (1911.9)	
(>4.5)	3128.1 (2291.0)	
Coffee consumption (cups/wk)		0.85
(0)	2989.5 (2298.0)	
(>0 – 3)	2896.4 (2230.0)	
(> 3 – 14)	2751.4 (1765.9)	
(> 14)	3184.9 (2565.9)	
Race		0.42
White	2999.0 (2184.8)	
Asian	2367.6 (1468.9)	
Other	3192.3 (3203.7)	
Second-hand smoke exposure		0.72
<1	2918.3 (2097.7)	
1-4 hours per week	3211.9 (2896.6)	
>5 hours per week	2871.2 (2090.5)	
Age		0.62
20-24	3226.5 (1999.7)	
25-29	3051.5 (2345.4)	
30-34	2848.3 (2397.0)	
35-39	2364.1 (1666.6)	
40-44	3092.5 (2510.2)	
45-50	2667.4 (2237.6)	
BMI		0.99
<18.5	3070.9 (2199.8)	
18.5-25	2981.5 (2077.0)	
>25-30	2931.3 (2435.1)	
>30	2811.5 (2551.9)	
Centre		0.40
Kingston	3022.1 (2094.3)	
Ottawa	2732.0 (2553.4)	

*p-value from F test regressing each variable on total MET-min/wk

4.2.3 Variable Conceptualization for Continuous Variables

Before examining the relationship between the continuous covariates and tHcy these variables were considered in two different representations. Variables were considered continuously and categorically. Variables were divided into categories based on quartiles except for BMI, which was categorized by clinically predefined cut points. These different conceptualizations were considered in order to best represent each variable in the multivariate analysis.

Table 4.7 presents the p-value from the F-statistic from the linear regression (each covariate being regressed on tHcy) for both representations of each variable. The representation with the lower p-value was carried through to the multivariate analysis. Cholesterol, LDL cholesterol, systolic blood pressure, diastolic blood pressure, alcohol consumption, residential THMs, vitamin B12 and coffee consumption were determined to be best represented as continuous variables, and BMI, age, triglycerides, HDL cholesterol, and folate were determined to be best represented as categorical variables.

| **Table 4.7** | Variable Conceptualization | | | |
| Variable | Continuous representation | | Categorical Representation | |
	F-Value	P-Value	F-Value	P-Value
BMI	0.04	0.8417	0.72	0.5409*
Age	0.01	0.9167	2.12	0.0646*
Cholesterol	0.37	0.5416*	0.28	0.8404
Triglycerides	0.03	0.8586	0.93	0.4292*
HDL Cholesterol	0.00	0.9467	1.09	0.3540*
LDL Cholesterol	0.46	0.4976*	0.23	0.8778
Folate	3.17	0.0765	9.58	<0.0001*
Systolic BP	5.63	0.0185*	2.26	0.0824
Diastolic BP	4.55	0.0339*	1.78	0.1512
Alcohol Consumption	10.30	0.0015*	1.83	0.1417
Residential THMs	1.74	0.1883*	1.18	0.3169
B12	13.27	0.0003*	3.47	0.0169
Coffee	9.13	0.0028*	3.33	0.0203

* best representation of variable

4.2.4 Covariate Associations with Homocysteine

The crude relationship between tHcy and covariates was examined by regressing the covariates individually on tHcy. The F-statistic from the linear regression was used to determine if the relationship between tHcy and each variable was statistically significant. The mean tHcy value, the parameter estimate (with confidence limits), and the p-values from the F-statistic are presented in table 4.8. For the categorical variables, a statistically significant relationship was observed between tHcy and sex, and serum folate status, and a weak relationship was observed between tHcy and age. For sex, the parameter estimate indicates that female's tHcy is on average 1.36μmol/L lower than male's tHcy. For the continuous variables, a statistically significant positive relationship was observed between tHcy and alcohol consumption, and coffee intake. A statistically significant inverse relationship was observed between tHcy and alcohol consumption, vitamin B12, and systolic blood pressure.

Table 4.8   Relationships between tHcy and Covariates

Variable	Mean tHcy	Parameter Estimate (95% CI)	P value*
Sex			< 0.01
Male	9.38	Referent	
Female	8.02	-1.36(-1.88, -0.86)	
Smoking			0.69
Never	8.53	Referent	
Past	8.28	-0.25 (-0.96,0.45)	
Current	8.29	-0.24 (-1.03,0.55)	
Race			0.23
White	8.52	Referent	
Asian	8.46	-0.07 (-0.94, 0.80)	
Other	7.63	-0.89 (-1.90,0.12)	
Second-hand smoke exposure			0.14
<1	8.34	Referent	
1-4 hours per week	9.01	0.67(-0.07,1.42)	
>5 hours per week	8.92	0.58(-0.50,1.67)	

* P value from F statistic when variable is regressed on tHcy

Table 4.8 con't   Relationships between tHcy and Covariates

Variable	Mean tHcy	Parameter Estimate (95% CI)	P value
BMI			0.54
<18.5	7.71	Referent	
18.5-25	8.54	0.84 (-0.89, 2.57)	
25-30	8.48	0.78 (-1.00, 2.55)	
>30	8.04	0.33 (-1.55, 2.20)	
Age			0.06
(<25)	8.25	Referent	
(≥ 25 <30)	8.68	0.43 (-0.27,1.14)	
(≥ 30 < 35)	7.96	-0.28 (-1.18,0.61)	
(≥ 35 < 40)	9.43	1.18 (0.29,2.06)	
(≥ 40 < 45)	8.22	-0.03 (-0.82,0.76)	
(≥ 45 ≤ 50)	8.35	0.10 (-0.75,0.95)	
Triglycerides			0.43
(≤ 0.54 mmol/L)	8.60	Referent	
(> 0.54 ≤ 0.72 mmol/L)	8.37	-0.23(-0.93,0.48)	
(>0.72 ≤ 1.04 mmol/L)	8.71	0.12(-0.59,0.82)	
(>1.04 mmol/L)	8.16	-0.44-1.14,0.27)	
Serum Folate			<0.01
(≤ 22.00 nmol/L)	9.53	Referent	
(> 22.00 ≤ 28.69 nmol/L)	8.46	-1.07(-1.74,-0.39)	
(> 28.69 ≤ 35.88 nmol/L)	8.01	-1.52(-2.20,-0.85)	
(> 35.88 nmol/L)	7.88	-1.65(-2.33,-0.98)	
HDL Cholesterol			0.35
(≤ 1.11 mm/L)	8.43	Referent	
(>1.11 ≤ 1.29 mm/L)	8,56	-0.63 (-1.33,0.08)	
(>1.29 ≤ 1.48 mm/L)	8.11	-0.18 (-0.89, 0.53)	
(>1.48 mm/L)	8.74	-0.30 (-1.01,0.41)	
Centre			<0.01
Kingston	8.65	-0.81 (-1.39,-0.23)	
Ottawa	7.83	Referent	
Cholesterol per 1 std dev		-0.08 (-0.33,0.17)	0.54
Alcohol Consumption/ cup/day		0.50 (0.19, 0.81)	<0.01
B12 per 1 std dev		-0.41 (-0.63,-0.19)	<0.01
Coffee Consumption/ cup/day		0.24 (0.08,0.40)	<0.01
Diastolic BP per 1 std dev		0.27 (0.02,0.52)	0.03
LDL Cholesterol per 1 std dev		-0.09 (-0.36,0.18)	0.50
Systolic Blood Pressure per 1 std dev		0.28 (0.05,0.53)	0.02
Residential THM per 1 std dev		-0.17 (-0.42,0.08)	0.19

4.2.5 Physical Activity Association with Homocysteine

The crude relationship between physical activity and tHcy was examined by regressing each representation of physical activity (continuous, continuous with quadratic term, IPAQ categories, quintiles, and dichotomously based on participation in vigorous activity) individually on tHcy. The quadratic term was examined to test the hypothesis of a U-shaped relationship between tHcy and physical activity. The parameter estimate and the p-value for each regression can be found in table 4.9. The results of these regressions revealed no statistically significant relationship between physical activity and tHcy. The representation of physical activity with the most significant relationship with tHcy (based on the P value) was the IPAQ categories.

Table 4.9   Crude Relationship between Physical Activity and tHcy

Continuous MET Score	Parameter Estimate (95% CI)	P value*
Total MET Score per Standard Deviation	0.01 (-0.56, 0.89)	0.95
Total MET Score per Standard Deviation +	0.26 (-0.67, 1.23)	0.47
(Total MET Score per Standard Deviation)^2	-0.27 (-0.92, 0.86)	0.46
IPAQ		0.25
Inactive	0.31 (-0.64,1.26)	
Minimally Active	-0.34(-0.87,0.18)	
Health Enhancing Physically Active	Referent	
QUINTILES		0.33
(<1173 MET min/wk)	Referent	
(1173-2034 MET min/wk)	0.14(-0.66,0.93)	
(2034-3079.5 MET min/wk)	0.79(-0.00,1.57)	
(3079.5–4613 MET min/wk)	0.19(-0.62,0.99)	
(> 4613)	0.29(-0.51,1.08)	
VIGOROUS		0.85
Vigorous	Referent	
No Vigorous	-0.06(-0.62,0.51)	

* p value from F test when Physical activity is regressed on tHcy

4.3 Confounder Selection

In order to create the most parsimonious model to control for potential confounders a backward stepwise elimination procedure was used. All potential confounders were included in a model predicting tHcy and the variable with the largest p-value was removed. This process was continued until all variables had p-values ≤ 0.20. Using this process the covariates left in the model were serum folate, cholesterol, LDL cholesterol, alcohol consumption, residential THMs, vitamin B12, coffee consumption, sex, and race. The p values (from the partial F-statistic) for each of these variables in a model predicting tHcy can be found in table 4.10. The overall F statistic for this model was 9,67 with a p value of <0.001.

Table 4.10   Covariates to be included in the model

Variable	p
Serum Folate	<0.01
Cholesterol	0.04
LDL cholesterol	0.01
Alcohol Consumption	0.13
Residential THMs	0.09
Vitamin B12	<0.01
Coffee Consumption	0.04
Sex	<0.01
Race	0.16

4.4 Main Analysis

The relationship of interest in this study (relationship between tHcy and physical activity) was examined using a multiple linear regression. The confounders identified by the backward elimination method were entered into the model with each representation of physical activity. Possible interactions were examined by the later inclusion of interaction terms, and two sensitivity analyses were performed to provide confidence in the methods used.

4.4.1 Full Model

Each representation of physical activity was entered into the model individually. The associated p-value from the F test for each representation in the model, as well as the change in parameter estimate, can be found in table 4.11.

Table 4.11   Physical Activity in Full Model

	P	Parameter Estimate (95% CI)*
Total MET Score Per Std Dev	0.95	-0.01 (-0.34,0.12)
Total MET Score per Std Dev +	0.57	-0.18 (-0.94, 0.39)
Total MET Score^2 per Std Dev	0.56	0.19 (-0.49, 0.83)
IPAQ	0.68	
Inactive		0.31 (-0.55,1.17)
Minimally Active		-0.06 (-0.51,0.39)
Health Enhancing Physically Active		Referent
METQUINT	0.20	
(<1173 MET min/wk)		Referent
(1173-2034 MET min/wk)		0.24(-0.43,0.91)
(2034-3079.5 MET min/wk)		0.69(0.01,1.37)
(3079.5–4613 MET min/wk)		-0.05(-0.75,0.65)
(> 4613)		0.20(-0.50,0.90)
VIGOROUS	0.92	
NO		0.03 (-0.47,0.53)
YES		Referent

* Adjusted for Serum Folate, Cholesterol, LDL Cholesterol, Alcohol Consumption, Residential THMs, Vitamin B12, Coffee Consumption, sex, race

The Parameter Estimate provided for Total MET Score Per Std Dev was found to be -0.01. This number describes the change in tHcy for one standard deviation change in total MET Score (-0.01µmol/L). It should be noted that the confidence interval for this estimate straddles zero indicating it is not statistically significant. Two parameter estimates are provided for the model including both Total MET Score Per Std Dev and (Total MET Score)[2] Per Std Dev these number denote the a curve that would best describe the relationship between tHcy and physical activity. The confidence intervals for both of these parameter overlap zero indicating they are not statistically significant. The parameter estimate for the categorical variables indicate the average difference between each category and the referent category. For instance the parameter estimate of 0.31 for Inactive indicates that on average individuals who fall in the IPAQ category of inactive have a tHcy concentration that is 0.31µmol/L higher than those in the referent category (Health Enhancing Physically Active). It should be noted again, that the confidence intervals for this parameter estimate overlap zero indicating that this number is not statistically different from zero.

4.4.2 Interaction Terms

In order to statistically examine the possibility of effect modification interaction terms were included in the final model (individually) with the IPAQ representation of physical activity for three variables: sex, serum folate, and age. Interaction terms were made into dichotomous variables, serum folate was based on the 25th percentile of all participants (low < 21.5µmol/L and high >21.5µmol/L) and age was dichotomized at 40 years old. The p-value and parameter estimate for each interaction term can be found in table 4.12

Table 4.12   Interaction Terms

LOW FOLATE		HIGH FOLATE	
Category	Parameter Estimate (95% CI)		Parameter Estimate (95% CI)
low activity	0.58(-0.74,1.89)	low activity	0.16(-1.03,1.36)
minimally active	-0.17(-1.16,0.82)	minimally active	-0.01(-0.51,0.49)
HEPA active	REF	HEPA active	Referent
p-value = 0.35			
< 40 years old		> 40 years old	
Category	Parameter Estimate (95% CI)		Parameter Estimate (95% CI)
low activity	0.93(-0.43,2.29)	low activity	0.01(-1.10,1.12)
minimally active	0.55(-0.31,1.42)	minimally active	-0.28(-0.81,0.25)
HEPA active	REF	HEPA active	Referent
p-value = 0.38			
Male		Female	
Category	Parameter Estimate (95% CI)		Parameter Estimate (95% CI)
low activity	-0.12 (-1.80,1.56)	low activity	0.49(-0.51,1.49)
minimally active	-0.32 (-1.10,0.46)	minimally active	0.08(-0.48,0.63)
HEPA active	REF	HEPA active	REF
p-value = 0.66			

* adjusted for folate, cholesterol, LDL cholesterol, alcohol consumption, residential THM, vitamin B12, coffee consumption, sex, and race

4.5 Sensitivity Analysis

4.5.1 Confounder Selection

Backward elimination was used to identify confounders to include in the final model in order to have the same model for all representations of physical activity, however, the change in estimate method of confounder selection is considered the gold standard as it is an evaluation of the effect of potential confounders on the exposure-outcome relationship. To evaluate the confounder selection method used (backward elimination) the change in estimate method was used to identify confounders for one representation of physical activity (IPAQ categories). The confounders identified as significant by this method were BMI, age, folate, alcohol consumption, center, residential THM, vitamin B12, coffee consumption, sex and smoking status. The parameter estimates for physical activity IPAQ categories for this model were compared to those found using the backward elimination procedure and are noted in table 4.13. The parameter estimates and confidence intervals were found to be very similar for the confounder selection methods, the change in estimate method produced confidence intervals spanning 2.45, and 1.19 whereas the backward elimination method had confidence intervals spanning 2.35 and 1.15. The similarity in the results found by the change in estimate method to those found by backward elimination indicates that the backward elimination method of confounder selection led to the creation of appropriate models to predict tHcy.

Table   4.13   Sensitivity Analysis – Confounder Selection

	Change in Estimate Method	Backward Elimination Method
	Parameter Estimate (95%CI)	Parameter Estimate (95%CI)
Low Active	0.38 (-0.50,1.26)	0.31(-0.55,1.17)
Minimally Active	-0.10(-0.55, 0.36)	-0.06(-0.51,0.39)
HEPA Active	Referent	Referent

4.5.2 Dichotomous Outcome

It is possible that having what is clinically considered high tHcy ($\geq$10 µmol/L) is different then having tHcy lower than that and only high tHcy truly indicates a breakdown in the methionine-homocysteine biosynthesis. If this is the case, treating tHcy as a continuous variable (as was done in this research) would lessen the ability to see this relationship because it is considering every increase in tHcy as equal. In order to test this possibility a sensitivity analysis was run dichotomizing tHcy at 10 µmol/L. The results of this logistic regression can be found in table 4.14. Although several large odds ratios were observed the confidence intervals are quite wide. These results are suggestive of effects that were not consistent with the underlying hypothesis of a U shaped curve.

Table 4.14 Full model with Dichotomous Outcome

	p	Odds Ratio
IPAQ	0.25	
Inactive		0.37 (0.07,2.04)
Minimally Active		0.50 (0.20,1.23)
HEPA		1
METQUINT	0.11	
1 (<1173 MET min/wk)		1
2 (1173-2034 MET min/wk)		4.55 (0.83, 25)
3 (2034-3079.5 MET min/wk)		6.25 (1.26, 33.33)
4 (3079.5–4613 MET min/wk)		8.33 (1.57, 41.67)
5 (> 4613)		2.86 (0.47, 16.67)

*adjusted for serum folate, total cholesterol, LDL blood pressure, residential THM, vitamin B12, sex and race

Chapter 5: Discussion

The purpose of this research was to examine the relationship between physical activity and total plasma homocysteine concentration with a larger goal of helping to understand the etiology of cancer and identify modifiable risk factors. This research was nested within a larger cross-sectional study of healthy volunteers recruited from centers in Ontario and Nova Scotia.

5.1 Physical Activity (MET-min/wk) Distribution

The International Physical Activity Questionnaire (IPAQ) was used to gather information about and quantify participants' physical activities. Participation in many different combinations of vigorous, moderate and walking activity would result in a score similar to the median MET-minutes/wk score found in this study (2553 MET-min/wk). Some examples are:

1) 3 days a week of vigorous activity for 1 hour 15 minutes, 0 days a week of moderate activity and 7 days a week of walking for ½ hour

2) 3 days a week of vigorous activity for 1 hour, 7 days a week of moderate activity for 15 minutes and 7 days a week for 30 minutes

The physical activity score in this study can be compared to that found in other studies using the same study instrument. The IPAQ was used for physical activity population surveillance in a study in the European Union, in which they observed a mean activity score for 16230 participants of 1440 MET-min/wk (Rutten, 2004). There are several possible reasons for the lower score in the European Union study compared to the current research; one is that the age group examined in the European study included participants from a larger age range with some older participants (15-65+). If that study is limited to participants aged 15-54, the median physical activity score increases to 1705.5 MET-min/wk. Another factor that may have affected

the mean physical activity score is that in the present study the population of participants was made up of many individuals from university and hospital environments, as both recruitment areas were on university campuses. This may have lead to the recruitment of individuals who were more educated and those of higher socio-economic status (SES) than the population at large. People who are more educated and have higher SES tend to be healthier and more active. This is seen in the European study when participants were grouped by gross household income and found that those in the lowest income bracket had a median activity score of 1205.04 MET-min/wk, while those in the highest income bracket had a median score of 1668.00 MET-min/wk (Rutten, 2004). The combination of these two factors, along with possible difference in lifestyles of Canadians compared to Europeans could explain the difference in mean activity score found. In a second European study the median physical activity score found for approximately 4800 participants using the short version of the IPAQ was 2970.00 MET-min/wk. This study was administered in the summer months (May – October) which may explain the higher median physical activity score found compared to the other European study and the current study.

The mean physical activity score found in the current study is, for the most part, consistent with those found in similar studies using the same study instrument. As well, the goal of the measurement of physical activity for this study was not to provide an accurate MET-min/wk score to each participant, rather it was to be able to rank individuals activity levels when compared to each other. This was completed with confidence due to the high reliability of the survey instrument and the consistency in which the survey was administered to all participants.

5.2 Total Plasma Homocysteine Concentration Distribution

The primary outcome of interest in this study was total plasma homocysteine (tHcy). This outcome was used as a biomarker of the methionine-homocysteine biosynthesis. A breakdown in this cycle can lead to several potential cancer causing mechanisms. For this research tHcy had a normal distribution ranging from 3.91μmol/L – 17.53μmol/L. Several other studies measuring the relationship between physical activity and homocysteine have had larger ranges of homocysteine values (ie 3.6 – 137 (Nygard 1995) and 3 – 132 (Ganji, 2003)). The limited range of homocysteine values found in this study could be due to several factors. This study included healthy adults age 20-50, the exclusion of older participants and of individuals with a history of outcomes that are known to increase homocysteine levels would have kept homocysteine levels lower than those found in previous studies. Also, Canada supplements cereal and grain products with folic acid. As has been discussed, folate has an inverse relationship with tHcy and this supplementation helps to keep homocysteine levels of people in Canada lower. These factors, that may have affected the distribution of tHcy in this study and made it different from those found in other studies, were all measured and controlled for in the final model.

5.3 Relationships

5.3.1 Relationships Between Covariates

Many statistically significant relationships were observed between the covariates. The relationships between the continuous covariates were examined by looking at correlation coefficients, those between continuous and categorical covariates were examined by looking at the p-values from f-statistics from simple regressions, and categorical-categorical relationships were examined through the use of chi-square tests. Some of the strongest relationships were

65

observed between triglycerides and cholesterol, systolic blood pressure and BMI, LDL cholesterol and overall cholesterol, LDL cholesterol and triglycerides, and diastolic and systolic blood pressure. These correlations were understandable and expected as these variable pairs measure similar physiologic parameters. As a result, only one variable from each of the above pairs was represented in the co-variate model.

Other relationships that were observed in this data are highlighted in table 4.4. Many of these relationships were expected and confirm that the data collected for this study follows trends found in previous research including increased coffee consumption being associated with increased blood pressure (both systolic and diastolic), older age and increased blood pressure, and older age and higher BMI.

5.3.2 Relationships Between Covariates and Physical Activity

A statistically significant relationship was observed between physical activity and sex and coffee consumption. The relationships observed indicated that increased coffee consumption was associated with increased physical activity and that females were more active than males.

5.3.3 Relationships Between Covariates and Homocysteine

Many of the expected relationships were observed between tHcy and the covariates measured. The strongest relationships observed were the inverse relationship between tHcy and vitamin B12 and serum folate, and the relationship between sex and tHcy (with males having higher tHcy then females). These relationships are consistent with those found in previous literature (Refsum, 2006). No relationship was observed between age and tHcy, where in most literature there is a positive relationship (Refsum, 2006). The reason no relationship was found in this study can likely to attributed to the limited age range of the study population. Homocysteine

levels generally start to rise between the ages of 40-60 and the maximum age for participation in this study was 50.

Other relationships that were observed in the largest study of life style determinants of tHcy to date (Nygard, 2002) include a positive relationship with systolic and diastolic blood pressure, coffee intake, smoking, and total cholesterol. Consistent with those results this research found positive relationships between diastolic and systolic blood pressure and coffee consumption respectively, with tHcy. However, no relationship was observed between tHcy and smoking status or total cholesterol. The former may be due to the limited number of smokers in this study (27) which limited the power to see this relationship.

5.3.4 Relationships Between Physical Activity and Homocysteine

The models in this study did not show a significant relationship between any of the representations of physical activity and homocysteine.

Of the studies that have examined this relationship in the past only three looked at a population with a similar age distribution, one of which only included people with increased risk of CVD at baseline (Chrysohoou, de Bree, Husemoen). Two of these studies were cross-sectional with mean age 42 and 41 respectively and the third was a controlled trial with individuals aged 30-60. All three of these studies found no association between physical activity and homocysteine. Studies that did find an association included people only 40 years of age or older (most were over age 60) (Mora 2006, Hellgren 2004, Danker 2004, Nygard 1995). This trend, along with evidence that as people get older homocysteine levels change, indicates that the biological mechanism that may explain a relationship between physical activity and homocysteine differs between a young healthy adult population and an older adult population.

67

When a stratified analysis was performed in this study based on age (group one <45 and group two ≥45) there was still no relationship observed between physical activity and homocysteine. This analysis had limited statistical power as 28 participants in this dataset were 45 years old or older.

Physical activity was represented total MET scores, dichotomously based on vigorous/no vigorous activity and by using the 3 IPAQ categories. None of these representations allowed for the discernment of individuals taking part in intense training. MET scores can be high through the participation in large amounts of moderate activity, the involvement in some vigorous activity does not indicate regular intense training, and the highest IPAQ category is activity that is thought to be 'health enhancing' and is not the same as intense training. Using the data collected there was no way to put the people who were involved in regular intense training into a separate category in order to look at the relationship between homocysteine and involvement in this type of physical activity.

5.4 Methodology

5.4.1 Outcome Assessment

The measurement of the outcome of interest (tHcy) was performed as has been outline by the American Society of Human Genetics (American Society of Human Genetics, 1998). The outcome is not subject to recall bias, or selection due to the use of carefully performed blood tests to ascertain this variable. Any blood vials sent to the testing centre that were identified as possibly being mislabelled were discarded.

Misclassification of the outcome is a possibility in two respects. The first is that for this research a single blood sample was used to ascertain tHcy and the validity of a single

measurement to represent an individual's tHcy can be questioned due to the possibility of variance based on time of day, week, or month. The variance based on time was limited as all blood samples were taken at the same time of day (in the morning between 8am and 10am). As well, two studies have been done looking at the within individual variability the results of both suggest that tHcy concentrations are stable, and that a single blood sample is a reliable measurement to represent this outcome (Garg, 1997 and Clarke 1998).

The second possibility for misclassification of the outcome comes from using tHcy as a proxy measurement for the true outcome of interest: the functionality of methionine-homocysteine biosynthesis. Elevated homocysteine is one indication of dysfunction of this cycle; however when there is a build-up of homocysteine within the cell, the reaction from SAH to homocysteine reverses in order to decrease homocysteine levels (leading to an increase in SAH levels). The reversibility of this reaction means that tHcy is not the most sensitive marker of dysfunction of the cycle. Only when homocysteine levels are high within the cell is the excess homocysteine released into the plasma leading to elevated tHcy. Homocysteine levels above a certain threshold within the cell causes the reaction to reverse, lowering levels again. A more sensitive marker of the cycle may be SAH levels or SAH/SAM ratio which would reveal if there is a build up of SAH within the cell. Data on these levels will be available upon completion of the larger study.

5.4.2 Exposure Assessment

The constraints of the main study on the type of exposure representation chosen for this research included the need for the instrument to be concise and feasible. The IPAQ MET those needs. As a recall-based self-report questionnaire it is inexpensive, has low participant burden, it

can be administered quickly and captures both quantitative and qualitative information. The objective of this study was to quantify individual's physical activity patterns and rank individuals relative to each other. The questionnaire provided a basic quantification of the major behavioural characteristics of the activity patterns reported. The use of a questionnaire meant that the measurement of physical activity was reliant on an individual's own account of their activity. As participants did not know their homocysteine concentrations when filling out their questionnaires, the potential for recall bias was avoided. It has been found, however, that while physical activity questionnaires have good validity for high intensity and sedentary activities (when compared with calorimeters, accelerometers, and maximal oxygen uptake), they have poor validity for low and moderate intensity activity (Jacobs, 1993).

As well, physical activity is a socially desirable behaviour (Warnecke, 1997) and studies have found that social desirability may influence self-reported physical activity on some survey instruments (Brown 2004). If this social desirability bias results in all participants over-reporting their physical activity to a similar degree there would be no effect on the results of this study when using the continuous physical activity measurement. For this study the absolute amount of physical activity is not of interest, rather the relative amount compared to other participants. If all participants' physical activity is skewed to the right the same amount their relative position to one and other does not change. This bias would however affect the dichotomous representations of physical activity, potentially causing misclassification.

If the social desirability bias resulted in sedentary people over-reporting their physical activity levels more than those who were active, this would lead to an apparent diminished variability in the exposure and skew the results toward the null. There are conflicting results in

70

the literature on how much social desirability affects the measure of physical activity when using a self-report questionnaire (Warenecke 1997, Motl 2004).

The main limitation of the questionnaire for the purpose of this study was that questions regarding physical activity are subjective and can be interpreted differently by different individuals. In an attempt to over come this subjectivity, examples were given of the different types of activity inquired about (moderate and vigorous). The questions in the questionnaire asked about broad ranges of activity (e.g. moderate activity) and these types of questions when compared to questions, about specific activities (i.e. "Did you run for exercise this week?") are more cognitively challenging and can lead to inaccurate reporting (Welk, 2002). A second limitation is that individuals were asked to recall their activity for an average week in the past month. Whenever people are asked to recall information regarding something that happened in the past there is a chance of error. That being said, this questionnaire has been found to have good reliability and validity (when compared to an accelerometer) (Craig et al 2003).

To improve on the measurement of physical activity in future studies a diary-based self-report instrument could be used. An activity record, which instructs participants to record the individual episodes of activity as they occur during the day would provide more detail about the types, intensity, and patterns of activity. This type of physical activity assessment would overcome any challenges that a participant would have in recalling their activity patterns and they would overcome misinterpretation of questions as the specific type, intensity and frequency of each activity would be recorded. This type of instrument however, would be much more time consuming for the participant (especially as this study is attempting to ascertain individuals

average involvement in physical activity so it would have to be over a time period of at least one week if not longer).

Another option for physical activity measurement would be the use of activity monitors. This type of instrument is an objective indicator of body movement, it provides information on intensity, frequency and duration of activity, it is non-invasive and provides for ease of data collection and analyses (Welk,2002). Unfortunately these instruments are expensive which prohibits the assessment of large numbers of participants, and they have been known to be inaccurate in the assessment of some activities.

5.4.3 Study Design

Another potential limitation is the cross-sectional nature of this study. This study examined the relationship between physical activity and tHcy at a single point in time (although physical activity data represented activity in the past month) and this can bring into question the ability to establish the temporal sequence of exposure and outcome. For the relationship in question (between physical activity and tHcy) this is less of a concern because it is unlikely tHcy would affect an individual's physical activity patterns is unlikely. It should be noted however, that B12 deficiency (which is linked with higher tHcy concentrations) can lead to anaemia. One of the symptoms of anaemia is fatigue which could plausibly lead to lower activity levels. This would not likely affect the results of this study as only 23 of the 231 participants or 10% of the study sample had what is considered low B12 (<132 picomols/L) (Gupta A, 2004) and mild B12 deficiency does not usually result in fatigue (only 6 participants had B12 <100picomols/L).

72

5.4.4 Assessment of Potential Confounders

One of the main strengths of this study was the accurate collection of potentially significant confounders identified in previous research. Refsum et al. states in a 2006 study that 'it remains to be shown whether regular physical activity directly influences total plasma homocysteine concentration or whether the inverse association observed in some studies reflects an overall healthier lifestyle' (Refsum 2006). By including a number of potential confounders that are associated with a healthy lifestyle this study isolated physical activity from the group of healthy behaviours that are often found together. The use of blood tests to control for potentially strong confounders (Folate and vitamin B12) provided an unbiased measure of these factors.

5.5 Generalizability

The majority of participants in this study were Caucasian, and it is possible that the association between physical activity and tHcy may differ by race as the frequency of genetic polymorphisms often differ by race and it is possible that a polymorphism could play a role in this relationship such as methylene-Tetra-Hydro-Folate-Reductase (MTHFR). For this reason the results of this study cannot be generalized to a people who are not Caucasian.

5.6 Sample Size

The intended sample size for this study was two hundred and fifty; due to time limitations the ultimate sample size was limited to two hundred thirty one. This decrease in sample size limited the power of the study to be able to reveal an association, if one exists, between physical activity and tHcy. The larger cross sectional study is still under way (with a total sample size of seven hundred and fifty) and once all data has been collected the new data will be added and calculations will be rerun to fully examine this relationship.

5.7 Resistance vs Endurance training

Three studies (one cross sectional and two randomized trials) have investigated the potential effect of resistance vs endurance training on tHcy (Chrysohoou 2004, Herrmann 2003, Vincent 2003) with conflicting results. One study looked at endurance activity compared to resistance activity or a sedentary lifestyle and found a protective association with those taking part in endurance activities having lower homocysteine levels (Chrysohoou 2004). While the other two studies found that homocysteine levels are increased in highly aerobic exercise trained athletes (Herrmann 2003) and that a resistance training program induced protective effects by decreasing tHcy (Vincent 2003). The study that found that resistance training induced protective effects was on a small population (43 participants) of older adults (aged 60-80 years) (Vincent 2003), and the study finding that endurance training increased tHcy was on a young population (age 14-18 years). The lack of consistency of these results and the limitations of the study designs indicates that more research is required to answer the question of what the effect type of physical activity (endurance vs resistance training) has on the relationship between physical activity and plasma homocysteine levels.

5.8 Conclusion

Elevated homocysteine levels have been found to be associated with increased risk of cancer at several sites, whereas physical activity has been found to reduce cancer risk. This study investigated the relationship between tHcy and physical activity with a hypothesis that elevated tHcy plays a role in mediating the physical activity-cancer relationship. This study found that there is no relationship between physical activity and tHcy in a healthy adult population, which

suggests that tHcy does not play a role in the relationship between physical activity and cancer risk.

5.7 Future Direction

Although this study did not reveal any association between physical activity and total plasma homocysteine, questions remain on this relationship. This study did not examine the difference between resistance and endurance training on physical activity, nor was it able to truly discern the intensity of activity.

References

Ainsworth BE, Haskell WL, Leon AS, Jacobs DR Jr, Montove HJ, Sallis JF, Paffenbarger RS Jr. Compendium of physical activities: classification of energy costs of human physical activities. Med Sci Sports Exerc. Jan:25(1):71-80, 1993.

American Society of Human Genetics/American College of Medical Genetics Test and Technology Transfer Committee Working Group. Measurement and Use of Total Plasma Homocysteine. Am J Hum Genet 63:1541-1543, 1998.

Aguilar B, Rojas J, Collados M. Metabolism of Homocysteine and its Relationship with Cardiovascular Disease. Journal of Thrombosis and Thrombolysis 18(2), 75-87, 2004.

Bailey DM, Davies B, Baker J. Training in hypoxia: modulation of Metabolic and cardiovascular risk factors in men. Med Sci Spors Exercise 32:1058-66, 2000.

Biesma R, Schouter L, Dirx M, Goldbohm A. Physical Activity and risk of Ovarian Cancer: Results from the Netherlans Cohort Study. Cancer Causes and Control 17(1): 109-115 2006.

Boreham, C A., Kennedy, R A., Murphy, M H., Tully, M., Wallace, W F M., Young, I., Training effects of short bouts of stair climbing on cardiorespiratory fitness, blood lipids, and homocysteine in sedentary young women. British Journal of Sports Medicine. 39(9):590-593, September 2005.

Boushey CJ, Beresford ASS, Omenn GS et al: A quantitative assessment of plasma homocysteine as a risk factor for vascular disease. JAMA 274:1049-1057, 1995.

Brown W. Bauman A. Chey T. Trost S. Mummery K. Comparison of Surveys Used to Measure Physical Activity. Australian and New Zealand Journal of Public Health. 28(2)(pp 128-134), 2004.

Caspersen, C. J., M. A. Pereira, and K. M. Curran. Changes in physical activity patterns in the United States, by sex and cross-sectional age. Med. Sci. Sports Exerc., 32, No. 9, pp. 1601-1609, 2000.

Caspersen CJ, Powel KE, Chritenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Reports. 100(2)pp126-131, 1985.

Choi S, Mason J. Folate Status: Effects on Pathways of Colorectal Carcinogensis. American Society for Nutritional Sciences. 132: 2413S-2418S, 2002.

Chrysohoou, C., Panagiotakos, D., Pitsavos, C., Zeimbekis, A., Zampelas, A., Papademetrio, L., Masoura, C., Stefanadis, C. The associations between smoking, physical activity, dietary habits and plasma homocysteine levels in cardiovascular disease-free people: the 'ATTICA' study. Vascular Medicine May;9(2):117-23, 2004.

Clarke R, Woodhouse P, Ulvik A, Frost C, Sherlinker P, Refsum H, Ueland P, Khaw K. Variability and determinants of total homocysteine concentrations in plasma in an elderly population. Clinical Chemisty 44:1 102-107, 1998.

Colditz G, Cannuscia C, Frazier C. Physical activity and reduced risk of colon cancer: implications for

prevention. Cancer Causes and Control 8(4): 649-667, 1997.

Craig C, Marshall A, Sjostrom M, Bauman A, Booth M, Ainsowrth B, Pratt M, Ekelund U, Yngve A, Sallis J, Oja P. International Physical Activity Questionnaire: 12-Country Reliability and Validity. Medicine and Science in Sports & Exercise 1381-1390, 2003.

Dankner, R., Cheitrit, A., Sela, B-A., Physical activity and lower homocysteine levels in an elderly population. Medicine & Science in Sports & Exercise. 34(5) Supplement 1:S122, May 2002.

de Bree A, Verschuren M, Henk B, Kromhout D. Lifestyle factors and plasma homocysteine concentrations in a general population sample. American Journal of Epidemiology Vol 154, No 2, 2001.

Duncan, G., Michael P, Anton S, Limacher M, Martin D, Lowenthal D, Arning E, Bottiglieri T, Stacpoole P. Effects of exercise on emerging and traditional cardiovascular risk factors. Preventive Medicine 39 894-902, 2004.

Durand P, Prost M, Loreau N, Lussier-Cacan S, Blache D. Impaired Homocysteine Metabolism and Atherothrombotic Disease Lab Invest 81:645-672, 2001.

Eichholzer M, Luthy J, Moser U, Fowler B. Folate and the risk of colorectal, breast and cervix cancer: the epidemiological evidence. Swiss Med Wkly 131:539-549, 2001.

Fenech M, Aitken C and Rinaldi. Carcinogenesis, 18 1329-1336, 1997.

Fenech M, Dreosti and Rinaldi. Carcinogenesis, 19 1163-1171, 1998.

Friedenreich C, Orenstein M. Physical Activity and Cancer Prevention: Etiologic Evidence and Biological Mechanisms. The Journal of Nutrition 3456S-3463S, 2002.

Garg U, Zheng Z, Folsom A, Moyer V, Tsai M, McGovern P. Short-term and long-term variability of plasma homocysteine measurement. Clinical Chemistry 43:1 141-145 1997.
Hauge W. Homocysteine and pregnancy. Best Practice and Research Clinical Obstetrics & Gynaecology. Vol 17, No 3 pp 459-469, 2003.

Gaume V., Mougin F., Figard H., Simon-Rigaud M.L., N'Guyen U.N., Callier J., Kantelip J.P., Berthelot A. Physical Training Decreases Total Plasma Homocysteine and Cysteine in Middle-Aged Subjects. Nutrition and metabolism 29:125-131, 2005.

Hauge W. Homocysteine and pregnancy. Best Practice and Research Clinical Obstetrics & Gynaecology Vol. 17, No. 3, pp. 459-469, 2003.

Hellgren, M., Melander, A., Ostgren, C.-J., Rastam, L., Lindblad, U. Inverse association between plasma homocysteine, sulphonylurea exposure and physical activity: a community-based sample of type 2 diabetes patients in the Skaraborg hypertension and diabetes project. Diabetes, Obesity & metabolism. 7(4):421-429, July 2005.

Herrmann M., Schorr, H., Obeid, R., Scharhag, J., Urhausen, A., Kindermann, W., Herrmann, W. Homocysteine Increases During Endurance Exercise. Clin Chem Lab Med 41(11):1518-1524, 2003.

Herrmann M., Wilkinson J., Schorr H. Comparison of the influence of volume-oriented training and high-intensity interval training on serum homocysteine and its cofactors in young, healthy swimmers. Clin Chem Lab Med 41:525-31, 2003.

Husemoen L, Thomsen T, Jorgensen T. Changes in lifestyle and total homocysteine in relation to MTHFR(C677T) genotype: the Inter99 study. European Journal of Clinical Nutrition 60, 614-622, 2006.

IPAQ Research Committee 2004. Guildelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ) – Short Form Version 2.0. 04/2004.

Jacobs DR, Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evluation of 10 commonly used physical activity questionnaires. Med Sci Sports Exerc 25:81-91, 1993.

Jacobsen D, Homocysteine and vitamins in cardiovascular disease. Clin. Chem. 44:1833–1843, 1998.

Kato, I., Dnistrian, AM., Schwartz, M., Toniolo, P., Koenig, K., Shore, RE., Akhmedkhanov, A., Zeleniuch-Jacquotte, A., Riboli, E. Serum folate, homocysteine and colorectal cancer risk in women: A nested case-control study. British Journal of Cancer. Vol. 79(11-12) (pp 1917-1922), 1999.

Kelsey J, Whittemore A, Thompson D. Methods in Observational Epidemiology. Oxford University Press, 1996.

Kleinbaum, Kupper, Muller. Applied Regression Analysis and Other Multivariable Methods. PWS-KENT Publishing Co 1988.

Kuo, H.-K., YEN, C.-J., Bean, J. F. Levels of homocysteine are inversely associated with cardiovascular fitness in women, but not in men: data from the National Health and Nutrition Examination Survey 1999-2002. Journal of Internal Medicine. 258(4):328-335, October 2005.

Laporte R, Montoye H, Caspersen C. Assessment of Physical Activity in Epidemiologic Research: Problems and Prospects. Public Health Reports March-April Vol. 100, No. 2, 1985.

Martinez M, Giovannucci E, Jiang R, Henning S, Jacobs E, Thompson P, Smith-Warner S, Alberts D. Folate Fortication, plasma folate, homocysteine and colorectal adenoma recurrence. Int. J. Cancer: 199, 1440-1446 (2006).

Martinez M, Henning S, Alberst D. Folate and colorectal neoplasia: relation between plasma and dietary markers of folate and adenoma recurrence. Am J Clin Nutr 2004;79:691-7.

Medina, M., Urdiales, J., Amores-Sanchez, M. Roles of homocysteine in cell metabolism. Eur. J. Biochem. 268, 3871-3882 , 2001.

Miller, Alan. The Methionine-Homocysteine Cycle and Its Effects on Cognitive Diseases. Alternative Medicine Review 8(1), 2003.

Mora S, Lee M., Buring J., Ridker P. Associatin of Physical Activity and Body Mass Index With Novel and Traditional Cardiovascular Biomarkers in Women. JAMA March, 2006 Vol 295. No 12.

Mudd H, Finkelstein J, Refsum H, Ueland P, Malinow M, Lentz S, Jacobsen D, Brattstrom L, Wilcken B, Wilcken D, Blom H, Stabler S, Allen R, Selhub J, Rosenberg I. Homocysteine and Its Disulfide Derivatives: A Suggested Consensus Terminology. Arterioscler. Thromb. Vasc. Biol. 20;1704-1706, 2000.

Nygard O, Vollset SE, resum H, Stensvold I, Tverdal A, Nordrehaug JE, Ueland PM, Kvale G. Total plasma homocysteine and cardiovascular risk profile. The Hordaland Homocysteine Study. Jama. 274:1526-33, 1995.

Oguma Y, Shinoda-Tagawa T. Physical Activity Decreases Cardiovascular Disease Risk in Women Review and meta-Analysis. Am J Prev Med 26(5), 2004.

Oikawa S, Katsuhika M and Kawanishi S. Oxidative damage to cellular and isolated DNA by homocysteine: implications for carcinogensis. Oncogene 22, 3530-3538, 2003..

Pogribny IP, et al. Alterations in hepatic p53 gene methylation patterns during tumor progression with folate/methyl deficiency in the rat. Cancer Lett 115(1):31-8, 1999.

Pols M, Peeters P, Kemper H, Grobbee D. Methodological aspects of physical activity assessment in epidemiological studies. European Journal of Epidemiology 14:63-70, 1998.

Quadrilatero J, Hoffman-Goetz L. Physical Activity and Colon Cancer. J Sports Med Phys Fitness 43:121-38, 2003.

Randeva HS, Lewandowski KC, Drzewoski J, et. al. Exercise decreases plasma total homocysteine in overweight young women with polycystic ovary syndrom. J.Clin Endocrinal METab 87:4496-501, 2002.

Refsum H, Nurk E, Smith A, Ueland P, Gjesdal C, Bjelland I, Tverdal A, Tell G, Nygard O, Vollset S. The Hordaland Homocysteine Study: A Community-Based Study of Homocysteine, Its Determinants, and Associations with Disease. American Society for Nutrition.V136(6 suppl)1731-1740, 2006.

Rothman K, Greenland S. Modern Epidemiology.Lippincott-Raven Publishers, 1998.

Rousseau A.S., Robin, S., Roussel, A.M., Ducros, V., Margaritis, I. Plasma homocysteine is related to folate intake but not training status. Nutrition, metabolism & Cardiovascular Diseases. 15:125-133, 2005.

Rundle, A. Moelcular Epdiemiology of Physical Activity and Cancer. Cancer Epidemiology, Biomakers and Prevention 2005;14(1). January 2005.

Sallis, J. Age-related decline in physical activity: a synthesis of human and animal studies. Medicine & Science in Sports & Exercise. 32(9):1598-1600, September 2000.

Sallis J, Saelens B. Assessment of Physical Activity by Self-Report: Status, Limitations, and Future Directions. Research Quarterly for Exercise and Sport; Jun; 71,2, 2000.

Samad A.K.A., Taylor R.S., Marshall T., Chapman M.A.S. A meta-anlysis of the association of physical activity with reduced risk of colorectal cancer. Colorectal Disease, 7, 204-213, 2004.

Semin Thromb Hemost. 26:263-79, 2000.

Selhub J. Homocysteine metabolism. Annu. Rev. Nutr. 19:217-46, 1999.

Sen, C. Oxidants and antioxidants in exercise. Anerucab Ogtsuikigucak Society 79(3): 675 – 686, 1995.

Todd R, Wong DT "Oncogenes". *Anticancer Res.* 19 (6A): 4729-46, 1999.

Vincent Kr, Braith RW, Bottiglieri T, Vincent HK, Lowenthal DT. Homocysteine and lipoprotein levels following resistance training in older adults. Prev Cardiol. 2003 fall; 6(4):197-203.

Wald D, Law M, Morris J. Homocysteine and cardiovascular disease:evidence on causality from a meta-analysis. BMJ Vol 325 Nov 23, 2002.

Weinstein S, Ziegler R, Selhub J, Fears T, Stickler H, Brinton L, Hamman R, Levine R, Mallin K, Stolley P. Elevated serum homocysteine levels and increased risk of invasive cervical cancer in US women. Cancer Causes and Contraol 12: 317-324, 2001.

Welk, G. Physical Activity Assessment for Health Related Research. Copyright 2002.

Williams, P. Physical fitness and activity as separate heart disease risk factors: a meta-analysis. *Med Sci Sports Exerc* 33:754-59, 2000.

Wright M, Francis K, Cornwell P. effect of acute exercise on plasma homocysteine. J Sports Med Phys Fitness 38:262-5, 1998.

Wu, L., Wu. J. Hyperhomocysteinemia is a risk factor for cancer and a new potential tumor marker. Elsevier Science. Acta 322 21-28, 2002.

 Yi P, Melnyk S, Pogribna M, Pogribny P, Hilne, James. Increase in plasma homocysteine associated with parallel increases in plasma S-adenosylhomocysteine and lymphocyte DNA hypomethylation J. Biol. Chem. 275: 29318-29323, 2000.
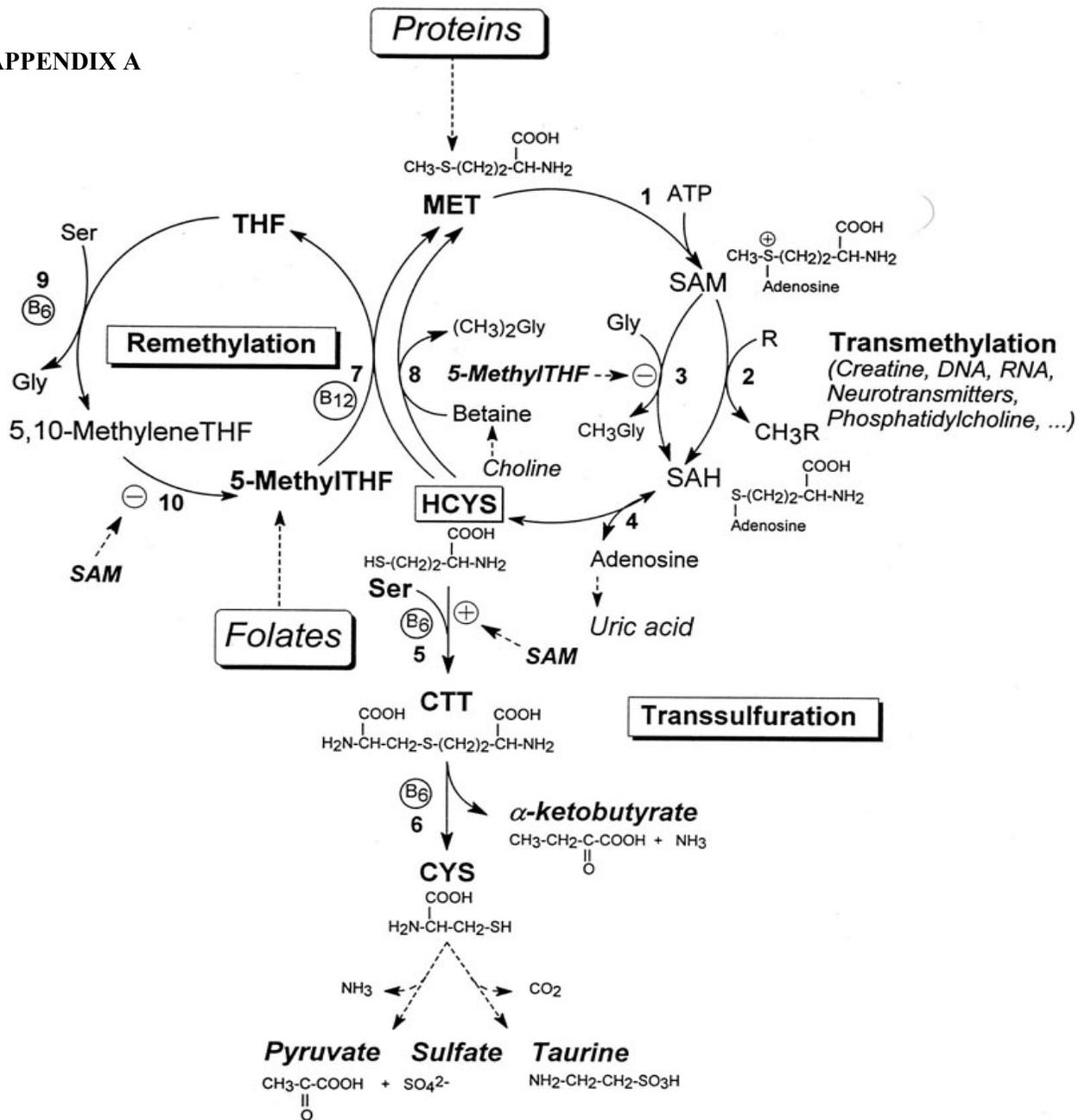
Fig A1 Detailed Methionine Metabolism (Durand et al, 2001)

(MET – Methionine, Gly – Glycine, Ser – Serine, CTT – cystathionine, CBS – cystathionin β-synthase, CYS – cysteine, HCYS – Homocysteine)

Appendix B Table A.1 Summary of Cross Sectional Studies

STUDY	N	POP/AGE	PA MEASURE	CONFOUNDERS	CONTRAST	RESULT
Nygard 1995	16176	Male and Female AGE: 40-67 EXCLUDE: - hypertension - diabetes - coronary heart disease - cerebrovascular disease * two age groups 40-42 and 65-67 (individuals age 43-64 years made up 2% of the study population)	Check one of four boxes Validated by the finding of a strong inverse relation between triglycerides & Physical Activity	- smoking - vitamin supp – (self report) -cholesterol and triglycerides (from blood sample) -NO FOLATE	Compared heavy training to sedentary activity groups	Inverse association in both age groups OR = -0.43 in younger age group when comparing heavy training to sedentary activity
Mora 2006	27158	Female AGE: 45 or older Apparently Healthy Health care professionals EXCLUDE: - Cardiovascular Disease - Cancer	Questionnaire - previous studies show validity and reliability Energy expended on PA/wk was calculated for each participant	- age - race - smoking - blood pressure - diabetes - menopausal status - hormone use - BMI - NO FOLATE	Quintiles of PA based on kcal/week found	Inverse association OR of 1.06 for lowest quintile of PA compared to highest quintile of activity
Hellgren 2004	387	Male and Female AGE: mean age 70 Patients with type 2 diabetes	Check one of four boxes	- age - sex - serum insulin - creatinine - LDL - cholesterol - hypertension - waist hip ratio - smoking - previous CDV - NO FOLATE	3 activity groups: no activity light moderate	Inverse association for both sexes
* Dankner 2004	423	Male and Female AGE: 62-75	Three questions - do you engage in activity? what kind? times/hours per week?	- smoking - B vitamin - BMI - blood pressure - kidney disease - ethnic origin - age - NO FOLATE	Sedentary vs Active participants	Inverse association between even small amounts of activity and tHcy in the elderly OR = 1.10 for sedentary vs active people
Chrysohoou 2004	2282	Male and Female AGE: mean age 42 EXCLUDE: - history of chronic disease - use of drugs that affect homocysteine levels (methotrexate, trimethoprin, cholestyrmine and cyclosporine)	Physical activity was defined as leisure-time activity of certain intensity and duration, at least once/wk during the past year Put in groups according to calories per min	- smoking - BMI - blood Pressure - cholesterol - age - alcohol - sex - coffee consumption - years of school - income - fruit and veg consumption - glucose - NO FOLATE	4 activity groups: light moderate vigorous inactive	No association
Rousseau 2005	82	Male AGE: mean age 27 74 well trained athletes 8 sedentary individuals EXCLUDE: - smokers - drinkers - BMI>35	7 day activity record - Appropriate MET value given to each participant and energy expenditure calculated	7 day food record - B6 - B12 - FOLATE	High energy expenditure compared to sedentary individuals	No association once folate was controlled for

		- family history of CHD - those treated by anti-inflammatory or any drug containing antioxidants				
de Bree 2001	3025	Male and Female AGE: mean age 41	PA was reported as the average amount of time spent in various activities over the previous year	- smoking - alcohol - coffee - tea - FOLATE	4 activity groups: sedentary moderate active very active	No association

Appendix C Table A.2 Summary of Randomized Control Trials

STUDY	N	POPULATION	TREATMENT GROUPS	RESULT
Herrmann 2003	20	AGE: 16 ± 2 years swimmers	3 weeks volume-oriented training and high intensity training	Positive association Homocysteine increased in both groups
Bailey 2000	14	Male Healthy, Caucasian, university students - No overt symptoms of poor health or disease Perform between 2-3 h of PA a week at intensity equivalent 120-140bpm EXCLUE: - smokers - taking any drugs	4wks of cycling exercise 3 times per week for 20-30 min at 70-85% of max heart rate	Positive association Homocysteine levels increased by 10%
Duncan 2004	324	Male and Female AGE: 48.9 ± 8.4 EXCLUE: -known chronic disease -engaged in structured PA	Randomized to four groups differing in training intensity (heart rate) and frequency (sessions per week) 6 months	Positive association * Homocysteine levels were within the normal range before and after training *
Husemoen 2006	915	Male and Female AGE: 30-60 assessed to be at increased CVD risk at baseline	Lifestyle intervention (high or low)	No association
Boreham 2005	15	Female AGE: mean age 18.8 recruited from university undergraduate population Subjects required to be sedentary at baseline EXCLUDE: - history of CVD - cigarette smoking - hypertension - obesity - musculoskeletal injury - taking pharmacotherapeutic drugs	control or stair climbing group 8wks progressive stair climbing programme	No association * Post hoc calculations suggest that the study only had sufficient power to detect a 1umol/l difference in homocysteine change between the intervention and control groups *
Vincent 2003	43	Male and Female AGE: 60-80	6 month high or low intensity resistance exercise or control group	Inverse association - Serum homocysteine decreased for the low and high intensity groups (by 5.30% and 5.34%

				respectively)
Randeva 2002	21 (12 exercisers, 9 non exercisers)	Female AGE: 30.6 $\pm$ 6.6 Overweight or obese with polycystic ovary syndrome	6 month exercise program	Inverse association Significant decrease in plasma total homocysteine

APPENDIX D

Categorical Investigation of Physical Activity (IPAQ Research Committee, 2004)

1. Inactive (CATEGORY 1)

This is the lowest level of physical activity. Those individuals who do not meet criteria for categories 2 or 3 are considered Inactive.

2. Minimally Active (CATEGORY 2)

The minimum pattern of activity to be classified as minimally active is any one of the following 3 criteria:

    a) 3 or more days of vigorous activity of at least 20 minutes per day OR
    b) 5 or more days of moderate-intensity activity or walking of at least 30 minutes per day OR
    c) 5 or more days of any combination of walking, moderate-intensity or vigorous intensity activities achieving a minimum of at least 600 MET – min/week

Individuals meeting at least one of the above criteria would be defined as achieving the minimum recommended to be considered 'minimally active'. This category is more than the minimum level of activity recommended for adults in current public health recommendations, but not enough for total physical activity when all domains are considered.

3. Health-Enhancing Physical Activity (HEPA active - CATEGORY 3)

'HEPA' active are those individuals who exceed the minimum public health physical activity recommendations, and are accumulating enough activity for a healthy lifestyle. There are two criteria for classification as 'HEPA active':

    a) vigorous-intensity activity on at least 3 days achieving a minimum of at least 1500 MET-minutes/week OR
    b) 7 or more days of any combination of walking, moderate-intensity or vigorous intensity activities achieving a minimum of a least 3000 MET-minutes/week.

PHYSICAL ACTIVITY

We are interested in finding out about the kinds of physical activity that you have done in a **usual week** over the **past month**.

Think about all the **vigorous** activities that you did in a usual week in the past month. **Vigorous** physical activities refer to activities that take hard physical effort and make you breathe much harder than normal. Think *only* about those physical activities that you did for at least 10 minutes at a time.

19. In a usual week in the past month, on how many days did you do **vigorous** physical activities
like heavy lifting, digging, aerobics, or fast bicycling?

   _____ days per week

   ☐ No vigorous physical activities ➡ Skip to question 21

20. How much time did you usually spend doing vigorous physical activities on one of those days?

   _____ hours per day

   _____ minutes per day

Think about all the **moderate** activities that you did in an average week in the past month. **Moderate** activities refer to activities that take moderate physical effort and make you breathe somewhat harder than normal. Think only about those physical activities that you did for at least 10 minutes at a time.

21. During a usual week in the past month, on how many days did you do **moderate** physical activities like carrying light loads, bicycling at a regular pace, or doubles tennis? Do not include walking.

   _____ days per week

   ☐ No moderate physical activities ➡ Skip to question 23

22. How much time did you usually spend doing moderate physical activities on one of those days?

   _____ hours per day

   _____ minutes per day

Think about the time you spent walking in a usual week in the past month. This includes at work and at home, walking to travel from place to place, and any other walking that you might do solely for recreation, sport, exercise, or leisure.

23. During a usual week in the past month, on how many days did you walk for at least 10 minutes at a time?

_____ days per week

☐ No walking ➡ Skip to question 25

24. How much time did you usually spend walking on one of those days?

_____ hours per day

_____ minutes per day

The last question is about the time you spent sitting on weekdays during a usual week in the past month. Include time spent at work, at home, while doing course work and during leisure time. This may include time spent sitting at a desk, visiting friends, reading, or sitting or lying down to watch television.

25. During a usual week in the past month, how much time did you spend sitting on a week day?

_____ hours per day

_____ minutes per day