VEHICULAR TRAFFIC EXPOSURE TO POLYCYCLIC AROMATIC HYDROCARBONS
AND BREAST CANCER RISK

Irina Mordukhovich

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in
partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department
of Epidemiology in the Gillings School of Global Public Health.

Chapel Hill
2013

Approved by:
Marilie Gammon
Amy Herring
Lawrence Engel
David Richardson
Regina Santella
ABSTRACT

Irina Mordukhovich: Vehicular Traffic Exposure to Polycyclic Aromatic Hydrocarbons and Breast Cancer Risk
(Under the direction of Marilie D. Gammon)

Polycyclic aromatic hydrocarbons (PAHs) are ubiquitous environmental pollutants, known human lung carcinogens, and potent mammary carcinogens in animal models. However, the association between PAHs and breast cancer in women is unclear. Vehicular traffic is a major source of ambient PAH exposure. This study evaluates the association between residential exposure to vehicular traffic-related PAHs and risk of breast cancer, overall and by tumor subtype, and within strata of nucleotide excision repair and base excision repair genotypes and fruit/vegetable intake. For this population-based study, residential histories, dietary intake, and other factors were assessed in 1996-1997 for 1,508 newly diagnosed breast cancer cases and 1,556 controls. Residential traffic exposure estimates were reconstructed using a validated model for the years 1960 through 1995. The following single nucleotide polymorphisms were genotyped: ERCC1 8092C/A, OGG1 Ser326Cys, XPA -4A/G, XPD Lys751Gln and Asp312Asn, XPF Arg415Gln, XPG Asp1104His, XRCC1 Arg194Trp and Arg399Gln. Medical records and archived tumor tissue were used to determine case tumor subtype.

In spline figures, which were used to inform quantile cutpoints for regression models, breast cancer risk was increased among women with the top 1% of traffic PAH exposures. Odds ratios (and 95% confidence intervals) for breast cancer, estimated using unconditional logistic regression, were modestly elevated for the top 5% of long-term 1960-1990 traffic PAH exposure...
estimates, compared with below the median (1.44 (0.78, 2.68)). Associations between recent traffic exposure in 1995 (top 5% vs. below the median) and breast cancer were attenuated toward the null (1.14 (0.80, 1.64)), but were stronger among women with low fruit/vegetable intake (1.46 (0.89, 2.40)) and hormone-receptor negative tumors (1.67 (0.91, 3.05)). Associations were approximately two- to three-fold stronger among women with variant alleles for \textit{XPD} (Lys751Gln) and \textit{XRCC1} (Arg194Trp), and wild-type alleles for \textit{XRCC1} (Arg399Gln) and \textit{OGG1} (Ser326Cys), when comparing the upper and lower tertiles of traffic exposure during 1995 or 1960-1990. This study reports positive associations between traffic-related PAH exposure and breast cancer risk among women with comparatively high long-term traffic exposures or among those with certain DNA repair genotypes, low fruit/vegetable intake or hormone receptor negative tumors, although confidence intervals were wide.
**TABLE OF CONTENTS**

List of Tables..............................................................................................................................................viii

List of Figures..................................................................................................................................................x

List of Abbreviations....................................................................................................................................xi

Chapter I: Background.................................................................................................................................1

  Breast Cancer Biology and Epidemiology.................................................................................................1

  Environmental Exposures and Breast Cancer.........................................................................................5

  PAH Exposure Sources and Associations with Breast Cancer.................................................................7

  Traffic and Air Pollution..............................................................................................................................28

  PAHs and Breast Cancer: Proposed Mechanisms....................................................................................46

  DNA Repair Mechanisms: Biology and Epidemiology............................................................................61

  Addressing Limitations of Previous Investigations.................................................................................76

  Summary and Conclusions.........................................................................................................................79

Chapter II: Research Methods....................................................................................................................81

  Research Aims, Hypotheses, and Study Rationale....................................................................................81

  Long Island Breast Cancer Study Project (LIBCSP)..................................................................................89

  Results from the LIBCSP: PAH-Related Exposures and Breast Cancer.................................................93

  PAH Exposure Assessment: A Historical Geographic Model.................................................................96

  Genotyping Assays....................................................................................................................................104

  Covariate Assessment...............................................................................................................................110

  Statistical Approach.................................................................................................................................112
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strengths and Limitations</td>
<td>120</td>
</tr>
<tr>
<td>Summary</td>
<td>125</td>
</tr>
<tr>
<td>Chapter III: Exposure to Traffic-Related Polycyclic Aromatic Hydrocarbons and Breast Cancer Risk</td>
<td>127</td>
</tr>
<tr>
<td>Introduction</td>
<td>127</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>128</td>
</tr>
<tr>
<td>Results</td>
<td>134</td>
</tr>
<tr>
<td>Discussion</td>
<td>136</td>
</tr>
<tr>
<td>Conclusions</td>
<td>139</td>
</tr>
<tr>
<td>Supplemental Material</td>
<td>145</td>
</tr>
<tr>
<td>Chapter IV: Polymorphisms in DNA Repair Genes, Traffic-Related Polycyclic Aromatic Hydrocarbon Exposure, and Breast Cancer Risk</td>
<td>148</td>
</tr>
<tr>
<td>Introduction</td>
<td>148</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>149</td>
</tr>
<tr>
<td>Results</td>
<td>154</td>
</tr>
<tr>
<td>Discussion</td>
<td>156</td>
</tr>
<tr>
<td>Conclusions</td>
<td>161</td>
</tr>
<tr>
<td>Chapter V: Discussion</td>
<td>167</td>
</tr>
<tr>
<td>Summary of Study Aims and Rationale</td>
<td>167</td>
</tr>
<tr>
<td>Summary of Results</td>
<td>167</td>
</tr>
<tr>
<td>Comparison to Previous Studies</td>
<td>168</td>
</tr>
<tr>
<td>Biological Plausibility and Potential Underlying Mechanisms</td>
<td>172</td>
</tr>
<tr>
<td>Study Limitations</td>
<td>173</td>
</tr>
<tr>
<td>Study Strengths</td>
<td>175</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 3.1. Distributions of cumulative residential vehicular traffic PAH exposure estimates...140
Table 3.2. Residential traffic polycyclic aromatic hydrocarbon exposure estimates by case-control status...141
Table 3.3. Associations between varying time ranges of exposure to PAHs from residential traffic and breast cancer risk...142
Table 3.4. Associations between residential vehicular traffic exposure and breast cancer risk, within strata of fruit/vegetable intake and menopausal status...143
Table 3.5. Associations between residential vehicular traffic PAH exposure and the risk of breast cancer categorized by tumor hormone receptor status...144
Table 3.6. Correlations between selected residential vehicular traffic PAH exposure variables...147
Table 4.1. Associations between traffic polycyclic aromatic hydrocarbon exposure in 1995 and breast cancer, stratified by DNA repair genotype...162
Table 4.2. Associations between traffic polycyclic aromatic hydrocarbon exposure in 1960-1990 and breast cancer, stratified by DNA repair genotype...164
Table 4.3. Associations between traffic PAH exposure and breast cancer, stratified by combined DNA repair genotypes in XPD, XRCC1, and OGG1...166
Table A2.1 Previously reported participant characteristics by case-control status...180
Table A2.2. Associations between traffic exposure in the year 1995 and breast cancer, overall and among long-term residents...182
Table A2.3. Associations between selected residential traffic exposure variables and breast cancer, removing participants with the highest exposure levels...183
Table A2.4. Associations between traffic polycyclic aromatic hydrocarbon exposure and breast cancer, varying the relative contribution of the pre-arrival surrogate...184
Table A2.5. Associations between traffic PAH exposure and breast cancer, turning off the intersection component of exposure model...185
Table A2.6. Associations between traffic PAH exposure and breast cancer, with partial imputation conducted using imputation by census place...186
Table A2.7. Associations between selected residential traffic exposure variables and breast cancer categorized by tumor characteristics……………………………..….187

Table A2.8. Previously reported main effects of DNA repair polymorphisms on breast cancer risk……………………………………………………………………188
LIST OF FIGURES

Figure A1.1 Metabolism of Polycyclic Aromatic Hydrocarbons in Relation to Breast Cancer Risk .................................................................179
LIST OF ABBREVIATIONS

ACS: American Cancer Society
Ah: Aryl hydrocarbon
AP: Apurinic
ATSDR: Agency for Toxic Substances and Disease Registry
BER: Base excision repair
BPDE: Benzo[a]pyrene diolepoxide
CA EPA: California Environmental Protection Agency
CARB: California Environmental Protection Agency Air Resources Board
CCA: complete case analysis
CI: Confidence interval
CK5/6: Cytokeratin 5/6
CO: Carbon monoxide
CYP: Cytochrome P450
EGFR: Epidermal growth factor receptor
EHII: Environment and Human Health, Inc.
ELISA: Enzyme-linked immunosorbent assay
ER: Estrogen receptor
ERCC1: Excision repair cross-complementing group 1
FEN1: Flap structure-specific endonuclease 1
FFQ: Food frequency questionnaire
FP: Fluorescence polarization
GIS: Geographic information system
GGR: Global genome repair
GST: Glutathione S-transferase
HER2: Human epidermal growth factor receptor 2
HR: Homologous recombination
HWE: Hardy-Weinberg equilibrium
IARC: International Agency for Research on Cancer
IQR: interquartile range
LIBCSP: Long Island Breast Cancer Study Project
LIG1: DNA ligase 1
MALDI-TOF: Matrix-assisted laser desorption/ionization time-of-flight
MI: Multiple imputation
MMR: Mismatch repair
NCRP: National Council on Radiation Protection and Measurement
NER: Nucleotide excision repair
NHEJ: Non-homologous end-joining repair
NTP: National Toxicology Program
OGG1: 8-oxoguanine DNA glycosylase 1
OR: Odds ratio
OSHA: Occupational Health and Safety Administration
PAH: Polycyclic aromatic hydrocarbons
PBS: Phosphate-buffered saline
PCNA: Proliferating cell nuclear antigen
PCR: Polymerase chain reaction
PM: Particulate matter
PR: Progesterone receptor
ROS: Reactive oxygen species
SD: Standard deviation
SNP: Single nucleotide polymorphism
TCR: Transcription-coupled repair
TSP: Total suspended particulates
US: United States
USDOT: United States Department of Transportation
US EPA: United States Environmental Protection Agency
UV: Ultraviolet
VOCs: Volatile organic compounds
WHO: World Health Organization
XP: Xeroderma pigmentosum
XPA, XPD, XPF, XPG: xeroderma pigmentosum groups A, D, F or G
XRCC1: X-ray repair cross complementing group 1
CHAPTER I: BACKGROUND

This investigation aims to examine associations between historical residential traffic-related polycyclic aromatic hydrocarbon (PAH) exposure and breast cancer, and to examine gene-environment interactions between traffic PAHs and DNA repair polymorphisms with respect to breast cancer risk. Secondary aims include examining effect modification of the traffic PAH-breast cancer association by fruit and vegetable intake and menopausal status, and evaluating associations between traffic PAHs and breast cancer with cases categorized according to tumor hormone receptor status and $p53$ mutation status.

The following chapter provides a detailed review of the relevant literature. Specifically, Chapter I presents information regarding breast cancer biology and epidemiology, PAH sources, including traffic and air pollution, and relevant metabolic pathways, including DNA repair processes and genetic polymorphisms.

BREAST CANCER BIOLOGY AND EPIDEMIOLOGY

Breast cancer is the most incident cancer, excluding non-melanoma skin cancers, and the second-leading cancer-related cause of death among women in the United States (US) (American Cancer Society [ACS] 2011). The American Cancer Society estimates that, in 2011, US age-adjusted breast cancer incidence and mortality were approximately 120.7 cases per 100,000 women and 24.0 deaths per 100,000 women, respectively. This translates to an estimated 230,480 new cases and 39,520 deaths in 2011 (ACS 2011). As of 2011, a US woman's lifetime risk for developing breast cancer was estimated to be 12.15% (ACS 2011). In contrast, male breast cancer is an uncommon disease (Weiss et al. 2005). This document focuses on female
breast cancer unless otherwise noted.

The human breast is primarily composed of lobules (glands which produce breast milk) and ducts for the release of milk (ACS 2009-2010). The breast also contains lymphatic, connective, and fatty tissues (ACS 2009-2010). Most diagnosed breast cancers are ductal or lobular adenocarcinomas (ACS 2009-2010, ACS 2011, Kelsey and Horn-Ross 1993), and are invasive (i.e. having spread out of one tissue type into another) rather than in situ (ACS 2009-2010, ACS 2011). Invasive ductal carcinoma is the most common form of the disease (ACS 2011).

The risk factors and etiology of breast cancer, both genetic and lifestyle/environmental, are incompletely understood. Confirmed genetic risk factors for breast cancer are divided into three categories: (1) rare, high-penetrance genetic mutations (e.g. in \textit{BRCA1} or \textit{BRCA2}) that confer a large personal risk of breast cancer, (2) rare, intermediate-penetrance genetic mutations (e.g. in \textit{ATM} or \textit{CHEK2}), and (3) low-penetrance, common single nucleotide polymorphisms (SNPs) that confer a low personal risk of breast cancer but are so prevalent as to be significant on a public health/population level (e.g. SNPs found in \textit{FGFR2} or \textit{LSP1}) (Turnbull and Rahman 2008).

Family history of breast cancer, which represents shared genetic and lifestyle/environmental factors (Lichtenstein et al. 2000), is also an established risk factor for the disease (ACS 2009-2010), especially for women with more than one affected relative (ACS 2009-2010). However, most women diagnosed with breast cancer do not have a family history of this disease (Mayo Clinic 2011).

Established non-genetic risk factors for breast cancer (ACS 2009-2010, ACS 2010, Hankinson et al. 2004) include age, reproductive history (late age at first birth, nulliparity or low
parity, late menopause, early menarche, and little or no lactation), medical history (exposure to ionizing radiation, hormone replacement therapy use, recent oral contraceptive use, high breast density, personal history of certain types of benign breast disease, and personal history of breast and certain other cancers), body mass index, alcohol intake, and low physical activity levels. Breast cancer risk factor profiles differ by menopausal status (Barlow et al. 2006, Velentgas and Daling 1994). For example, obesity is an established risk factor for postmenopausal breast cancer, but is negatively associated with premenopausal breast cancer risk (Rose and Vona-Davis 2010, Velentgas and Daling 1994).

Many breast cancer risk factors are related to systemic circulating estrogen levels (Brody and Rudel 2003, Pike et al. 1993). For example, in postmenopausal women, endogenous estrogen mainly originates in adipose tissue stores. Therefore, increased body size is likely to be related to postmenopausal breast cancer risk through estrogenic pathways (Hankinson 2005-2006). Circulating estrogen levels are consistently associated with postmenopausal breast cancer risk in epidemiologic research (Hankinson et al. 2004, 2005-2006, Kaaks et al. 2005, Key et al. 2002). Clinical investigations report that estrogen receptor antagonists, such as the drug tamoxifen, can prevent breast cancer development (ACS 2009-2010, Visvanathan et al. 2009).

Breast cancer risk factor profiles, both genetic and lifestyle/environmental, differ with respect to tumor estrogen and progesterone receptor (ER/PR) status (Althuis et al. 2004, Chen and Colditz 2007, Colditz et al. 2004, Garcia-Closas and Chanock 2008). Estrogen plays a role in the development of a subset of breast tumors: "ER+ tumors are those with the capacity to be stimulated by estrogen (Dickson and Stancel 2000) and among these, PR expression is considered indicative of an intact ER signaling pathway (Lapidus et al. 1998)" (Hankinson et al. 2005-2006). Inconsistent findings regarding potential risk factors in the breast cancer literature
could be attributable in part to differences in tumor hormone receptor status distribution between study populations (Chen and Colditz 2007).

Breast cancer risk factor patterns may also differ with respect to tumor subtypes defined by hormone receptor status in combination with other markers such as human epidermal growth factor receptor 2 (HER2/neu) or cytokeratin immunohistochemical expression status (Chen and Colditz 2007, Millikan et al. 2008, Susan G. Komen for the Cure 2010, Yang et al. 2011). More specifically, breast tumors are generally classified into the following molecular subtypes: luminal A (ER+ or PR+, HER2-), luminal B (ER+ or PR+, HER2+), HER2 type (ER-, PR-, HER2+), and triple-negative (ER-, PR-, HER2-). Using data on immunohistochemical expression of cytokeratin 5/6 (CK5/6) and epidermal growth factor receptor (EGFR) proteins, triple-negative breast cancer can be further categorized into basal-like (ER-, PR-, HER2-, CK5/6+ or EGFR+) and unclassified (ER-, PR-, HER2-, CK5/6-, EGFR-) subtypes (Chen and Colditz 2007, McCullough L. et al, unpublished data, Millikan et al. 2008, Susan G. Komen for the Cure 2010, Yang et al. 2011). Although it is too early to make definitive conclusions, research suggests that triple-negative breast cancer differs in etiology from hormonally responsive breast cancer, with studies reporting positive associations between triple-negative cancer, parity and premenopausal body mass (McCullough L et al, unpublished data, Millikan et al. 2008, Susan G. Komen for the Cure 2010, Yang et al. 2011). Triple-negative breast cancer comprises approximately 10-25% of all breast cancers, and is overrepresented among premenopausal and African American women (Chen and Colditz 2007, Carey et al. 2006, 2010, McCullough L et al, unpublished data). Hormonally responsive, luminal tumors are the predominant breast cancer subtypes in the United States, particularly among White, postmenopausal women (Potter et al. 1995).

Breast cancer subtypes also differ in prognosis (Dunnwald et al. 2007, Fan et al. 2006,
Pharoah et al. 1999, Ross and Fletcher 1998, Susan G. Komen for the Cure 2010). Triple-negative and HER2 type cases generally have worse prognoses than luminal cases (Carey et al. 2006, 2010, McCullough L et al, unpublished data, Susan G. Komen for the Cure 2010), and luminal A cases have the best prognosis of any subtype (Susan G. Komen for the Cure 2010). Cases with \(p53\) mutation-positive tumors tend to have worse prognoses than \(p53\) mutation-negative cases (Pharoah et al. 1999); \(p53\)-mutation positive breast tumors are more likely to be hormone-receptor negative (Olivier et al. 2006).

ENVIRONMENTAL EXPOSURES AND BREAST CANCER

Studies of immigrant groups report that breast cancer rates gradually increase from the levels of immigrants' country of origin to the levels of their new country. This process continues into subsequent generations (Buell 1973, Kelsey and Horn-Ross 1993, Kliwerer and Smith 1995). These results suggest the importance of lifestyle or environmental factors to breast carcinogenesis (Kelsey and Horn-Ross 1993). Also pointing to a lifestyle/environmental component to breast cancer development is the relatively low breast cancer concordance observed in studies of monozygotic twins (Brody and Rudel 2003, Lichtenstein et al. 2000, Mack et al. 2002).

in known lifestyle, reproductive, and demographic breast cancer risk factors (Laden et al. 1997, Laden and Hunter 1998, Sturgeon et al. 1995, 2003). However, differing environmental exposure profiles between regions may also play a role.

With the exception, to some extent, of accidents in which large amounts of radiation or environmental pollutants are released (Brody and Rudel 2003, Laden and Hunter 1998), investigations examining environmental exposures in relation to cancer outcomes are challenging to conduct given a number of methodological issues (Laden and Hunter 1998). These include (1) difficulty finding a truly unexposed group given the ubiquity of many environmental contaminants, (2) difficulty measuring exposures and reconstructing individual exposure histories, and (3) small hypothesized relative risks, requiring very large studies for sufficient statistical power (Laden and Hunter 1998).

Certain environmental chemicals have been hypothesized to increase breast cancer risk due to their genotoxic or hormonally active properties (Brody and Rudel 2003, el-Bayoumy et al. 1992, Laden and Hunter 1998, Morris and Seifter 1992, Nasca and Pastides 2008). However, most of these environmental contaminants have not shown a consistent association with breast cancer risk in epidemiologic research. For example, the potential relationship between organochlorines and breast cancer has been extensively studied (Nasca and Pastides 2008). The majority of investigations report null associations (Nasca and Pastides 2008), including in three large prospective cohort studies (Engel et al. 2005, Hunter et al. 1997, Reynolds et al. 2004), the largest case-control study on this topic to date (Gammon et al. 2002a), and a meta-analysis of 22 investigations (Lopez-Cervantes et al. 2004) (Nasca and Pastides 2008). Similarly, despite engendering speculation in both scientific and lay communities, exposure to electromagnetic fields was not associated with female breast cancer risk in well-powered epidemiologic
investigations (Gammon et al. 1998, Hulka et al. 2008). Studies may point to a weak association

High-level ionizing radiation is an established environmental risk factor for breast cancer
(Laden and Hunter 1998), as seen in studies of nuclear fallout and medical treatments (ACS
2009-2010, Laden and Hunter 1998, Land et al. 2003). Electric light at night, which affects
circadian rhythm and melatonin production (Stevens 2009a), is also consistently related to breast
cancer based on studies of sleep duration, blindness, non-day shift work, and light pollution
(ACS 2009-2010, Kolstad et al. 2008, Stevens 2009a,b) (Straif et al. 2007). Bulky PAH-related
DNA adducts, a form of DNA damage induced by polycyclic aromatic hydrocarbons (PAHs), are
the only other environmentally-related factor consistently associated with breast cancer risk in
epidemiologic research (Gammon et al. 2004a, Gammon and Santella 2008, Li et al. 1996, 2002,

PAH EXPOSURE SOURCES AND ASSOCIATIONS WITH BREAST CANCER

PAHs are ubiquitous environmental contaminants which contain at least two conjoined
These chemicals are formed as incomplete combustion by-products of organic matter (e.g. fossil
fuels; Phillips 1999, Samanta et al. 2002). Humans are exposed to PAHs via inhalation,
ingestion, or absorption through the skin (Morris and Seifter 1992). PAH exposure sources, such
as traffic, cigarette smoke, and grilled or smoked foods, differ with respect to their PAH profiles
(i.e. which specific PAHs are present) and contain varying combinations of known or suspected
non-PAH carcinogens (Bostrom et al. 2002, IARC 2010, Naumova et al. 2002). Individual PAHs
differ in structure and therefore in reactivity and carcinogenicity (Cavalieri et al. 1991, Leavitt et
PAHs are established mammary carcinogens in laboratory rodents (Hecht et al. 2002). Certain PAHs and PAH sources (e.g. vehicular exhaust, cigarette smoke) are classified as confirmed, probable, or possible human carcinogens by several health and environmental agencies, largely based on lung cancer studies (California Environmental Protection Agency [CA EPA] 1999, IARC 2010, Minnesota Department of Health 2004, National Toxicology Program [NTP] 2011, Straif et al. 2005, United States Environmental Protection Agency [US EPA] 2011). The association between PAHs and breast cancer in humans is not well researched and is therefore currently unclear (Gammon and Santella 2008).

Associations between PAHs and cancer risk are difficult to evaluate for several reasons. PAH sources are usually complex mixtures of various PAHs and non-PAH carcinogens. Also, exposure to PAHs is ubiquitous, which complicates defining a referent group (Boffetta et al. 1997). Associations between PAHs and cancer risk are likely to be influenced by individual susceptibility factors as well as PAH exposure levels, which both differ between study populations (Bostrom et al. 2002)

*Occupational PAH Exposure*

Occupational exposure to PAHs occurs among firefighters (Caux et al. 2002), traffic policemen (Merlo et al. 1998), toll booth operators (Tsai et al. 2004a), professional drivers (Boffetta et al. 1997), heavy equipment and crane operators, and road, railroad, dock, coke-oven, aluminum, and iron and steel foundry workers (Boffetta et al. 1997, Gammon and Santella 2008, Mastrangelo et al. 1996). Other occupational PAH sources include but are not limited to rubber production, shale oil extraction, wood impregnation, energy generation from coal, oil and other fuels (including coal gasification), carbon black, carbon electrode and calcium carbide production, coal tar distillation, roofing, chimney sweeping, and motor vehicle repair (Boffetta et
Certain workers are exposed to especially high levels of vehicular exhaust, a major PAH source. Relevant occupations include “transportation and garage work, underground mining, vehicle maintenance and examination, traffic control, logging, firefighting and heavy equipment operation” (IARC 1989).

Ambient levels of benzo[a]pyrene, a commonly measured carcinogenic PAH, range widely across occupational settings (i.e. between approximately 0.1 and 100,000 ng/m$^3$) (Angerer et al. 1997, Castaño-Vinyals et al. 2004, IARC 2010, Ovrebo et al. 1995). Occupational PAH exposure occurs mainly through inhalation of both gaseous and particulate PAHs and skin absorption (Boffetta et al. 1997, IARC 2010). PAHs may be absorbed through the skin from contact with coal or petroleum-derived products such as tar, soot, or pitch (Bostrom et al. 2002, IARC 2010).

Few investigations examining associations between PAHs and breast cancer are conducted in occupational settings (Cantor et al. 1995, Labreche et al. 2010, Petralia et al. 1999, Weiderpass et al. 1999). This is largely because of the paucity of women available for such research (Brody and Rudel 2003). Two studies did not find elevated breast cancer rates among women with higher estimated occupational PAH exposures (Cantor et al. 1995, Weiderpass et al. 1999). Another study reported a positive association between employment in motor vehicle repair and breast cancer risk (Band et al. 2000). One investigation reported increased premenopausal breast cancer among women exposed to PAHs based on occupational histories (OR = 1.82, 95% CI: 1.02, 3.16) (Petralia et al. 1999). Upon stratification by tumor estrogen receptor status, this association was limited to ER-positive tumors (OR = 2.27, 95% CI: 1.14, 4.54; Petralia et al. 1999). However, the effects of PAH independent of benzene could not be evaluated due to insufficient sample size in the relevant subgroup (Petralia et al. 1999). Finally,
a recent study reported associations between postmenopausal breast cancer and PAH exposure
determined from occupational histories, especially from petroleum sources and for exposure
prior to age 36 (Labreche et al. 2010). Upon stratification by tumor hormone receptor status,
associations with PAHs from petroleum sources were present only for ER and PR-positive breast
tumors. No significant or marginal findings were reported with respect to hormone receptor
status when evaluating occupational PAH exposure from all sources (Labreche et al. 2010).

General Population PAH Exposure: An Overview

In the general population, major PAH exposure sources include cigarette smoke
(Besaratinia et al. 2002), grilled, smoked or broiled foods (Phillips 1999), wood and coal-burning
stoves, house dust (Lewis 1999), PAH-contaminated food crops (Bostrom et al. 2002, Morris and
Seifter 2002, Phillips 1999), and air pollution from industrial emissions, traffic, and heating
PAH sources include cooking, candles and incense (Wallace 2000), contaminated soil (IARC
2010), certain pharmaceutical products (IARC 2010), space heaters, and wood, leaf and garbage
2004).

In the US, wood smoke is the largest source of ambient PAHs overall (Bostrom et al.
2002). However, traffic pollution is an important source of both indoor and outdoor ambient
PAHs. This is especially true in or near urban areas, where traffic is often the largest ambient
PAH source (Bostrom et al. 2002, Dubowsky et al. 1999, Dunbar et al. 2001). PAHs emitted
from wood burning show relatively little carcinogenic activity, in contrast to PAHs in vehicular
1990). Wood burning is also more likely to occur away from areas of high population density.
PAHs have been detected in water supplies, especially in urban areas (Bostrom et al. 2002, Morris and Seifter 1992, IARC 2010). This is due to traffic runoff (Ramesh et al. 2004), atmospheric deposition, and industrial pollution (IARC 2010, Ramesh et al. 2004). However, drinking water PAH levels are generally low due to the relative insolubility of PAHs and effective water treatment procedures (IARC 2010). Emissions from natural sources such as forest fires or volcanoes comprise a small fraction of PAHs relative to anthropogenic sources of these chemicals at ground level (Bostrom et al. 2002). It is estimated that, on average, an individual's exposure to carcinogenic PAHs from all sources amounts to approximately 3 μg per day in the general population (Castaño-Vinyals et al. 2004). Due to the ubiquity of PAH exposure, nearly all members of the general population have measurable concentrations of urinary PAH metabolites (Huang et al. 2004, IARC 2010).

Diet is the main source of PAH exposure among non-occupationally exposed non-smokers (IARC 2010, Phillips 1999, Ramesh et al. 2004). Average daily dietary PAH intake is estimated to range from several nanograms to several micrograms (IARC 2010). PAHs are found in a wide variety of foods, most notably barbecued, broiled, grilled or smoked meat. They are also found in other cooked foods, foods that are processed or preserved in certain ways (e.g. oils, cereals, breads), and in environmentally contaminated foods, including vegetables, dairy products, and seafood (IARC 2010, Larsson et al. 1983, Lijinsky 1991, Ramesh et al. 2004, Roth et al. 1998). Atmospheric PAH deposition leads to contamination of foods via water, soil and air (Bostrom et al. 2002, IARC 2010, Ramesh et al. 2004, Shabad and Cohan 1972). PAH deposition onto soil is proportional to ambient PAH concentrations, and crops grown near roads, industrial sites, or urban areas show higher levels of PAH dust than crops from less polluted
areas (Bostrom et al. 2002, Ramesh et al. 2004). Crop deposition is an important source of PAH exposure in the general population (Bostrom et al. 2002).

PAH dose from inhalation has been estimated to be approximately 5-10% of dietary dose in the general population (Bostrom et al. 2002, Lioy et al. 1988). One study reported that for male non-smokers between 19 and 50 years of age, the contributions of dietary and ambient sources to total PAH dose were 96.2% and 1.6%, respectively (IARC 2010, Menzie et al. 1992). However, environmental ambient PAHs can show stronger associations with PAH-DNA adducts than diet, smoking, or occupational exposures (Beyea et al. 2006, Eder 1999). It is known that “the carcinogenic risk of PAH mixtures is highly dependent on the exposure pathway” (Ramesh et al. 2004). The impact of different exposure routes on effective biological PAH dose is unclear (Bostrom et al. 2002).

Associations between PAH-related sources and breast cancer have been explored in general population settings. Briefly, surrogates for PAH exposure that have been evaluated in relation to breast cancer risk include cigarette smoke, grilled or smoked meat intake, PAH-DNA adducts, and air pollution exposure (Bonner et al. 2005, Gammon et al. 2004a, Rundle et al. 2000a, Steck et al. 2007, Terry and Rohan 2002). The results of these efforts are described in more detail in later sections of this document. In addition, one ecological study found a null association between very low-level PAH water contamination and breast cancer after adjustment for known breast cancer risk factors (Dean et al. 1988, Laden and Hunter 1998). Another ecological investigation reported positive associations between breast cancer and creosote-contaminated water (Dusich et al. 1980). No other identified studies have evaluated associations between water PAHs and breast cancer risk. Finally, one investigation reported null associations between urinary PAH markers (1-hydroxypyrene and 2-naphthol, but not benzo[a]pyrene
metabolites) and breast cancer risk (Lee et al. 2010).

**Tobacco Smoke**

**PAH Content of Cigarette Smoke**

Tobacco smoke contains several PAHs, including benzo[a]pyrene and dibenzo[a,l]pyrene, which are both potent mammary carcinogens in laboratory animals (Cavalieri et al. 1991, el-Bayoumy et al. 1995, Hecht et al. 2002). Other PAHs, such as dibenz[a,h]anthracene and benz[a]anthracene, are also found in cigarette smoke (Smith et al. 2000). Many non-PAH carcinogenic chemicals, including aromatic amines and N-nitrosamines, are found in tobacco smoke as well (Terry and Rohan 2002).

The concentration and carcinogenic activity of benzo[a]pyrene and other PAHs may be higher in sidestream than in mainstream smoke (IARC 2010, Jinot and Bayard 1996, Laden and Hunter 1998, Morris and Seifter 1992, Nasca and Pastides 2008, Nelson 2001). For example, a study of smokers found that benzo[a]pyrene exposure per 100 cigarettes ranged from 0.5 to 7.8 mg in mainstream smoke and from 2.5 to 19.9 mg in sidestream smoke (Castaño-Vinyals et al. 2004, IARC 1983).

**Biological Relevance of PAHs in Cigarette Smoke to Breast Cancer Risk**

Active smokers have increased urinary levels of the PAH metabolite, 1-hydroxypyrene (IARC 2010, Terry and Rohan 2002). Levels of this metabolite are also associated with second-hand smoking (Gunier et al. 2006). Nicotine and cotinine (a metabolite of nicotine) from cigarette smoke are found in breast fluid (Hecht et al. 2002, Petrakis et al. 1978), and smoking has been associated with PAH-DNA adduct levels in breast tissue (Hecht et al. 2002, Li et al. 1996, Perera et al. 1995). Therefore, it is known that metabolites and chemicals from cigarette smoke reach the breast (Hecht et al. 2002). Both active and passive smoking induce DNA
damage according to in vivo and in vitro studies as well as epidemiological investigations (Husgafvel-Pursiainen 2004, Lodovici and Bigagli 2009).

It is also known that certain PAHs, including those found in cigarette smoke (Hecht 2002), are highly carcinogenic to the mammary gland in laboratory animals (el-Bayoumy et al. 1995, Huggins and Yang 1962, Ranadive and Karande 1963, Santodonato 1997). Benzo[a]pyrene from cigarette smoke likewise induces neoplastic transformation of human breast epithelial cells in vitro (Russo et al. 2002).

Smoking can impact ovarian function, leading to earlier menopause onset (Kaufman et al. 1980, Tanko and Christiansen 2004). Cigarette smoke may also exhibit antiestrogenic properties, independent of effects on the ovaries (Band et al. 2002, Tanko and Christiansen 2004). For example, studies report that smoking induces hydroxylation of estrogens (specifically, via the 2-hydroxylation pathway); the metabolic products, 2-hydroxyestrogens, show little estrogenic activity and are rapidly cleared from the bloodstream (Meek and Finch 1999, Michnovicz et al. 1986, Tanko and Christiansen 2004). Similarly, an in vitro investigation using breast cancer cells reported that PAHs, a component of tobacco smoke, “inhibited estradiol-induced cell proliferation” (Santodonato 1997) (Chaloupka et al. 1992). The postulated effects of cigarette smoke or its constituents on ovarian function and estrogenic activity are supported by in vitro and in vivo experimental studies (Santodonato 1997, Tanko and Christiansen 2004), and by epidemiologic associations between active smoking and, for example, osteoporosis (Lane 2006) and decreased serum and urinary estrogen levels among premenopausal women (Tanko and Christiansen 2004). The resulting reduction in estrogen exposure could counteract the genotoxic effects of smoking, as circulating estrogen levels are consistently associated with breast cancer risk in epidemiological studies (Hankinson et al. 2004, 2005-2006). It should be noted that
PAHs also exhibit weak estrogenic activity (Santodonato 1997), and cigarette smoke has been reported to induce transcription via the estrogen receptor (Meek and Finch 1999).

The literature regarding passive smoking and endpoints related to ovarian function or estrogenic activity is relatively sparse. A recent investigation among young women reported that passive smokers had higher serum estrogen levels (for estrone, estradiol, estriol, and 16-hydroxyestrone) than active smokers, when smoking status was classified according to serum cotinine levels (Soldin et al. 2011). Serum estradiol and estriol levels were higher among passive smokers relative to both active smokers and non-smokers in this study, whether classified by serum cotinine levels or by self-report (Soldin et al. 2011). Another investigation, which did not include active smokers, reported negative associations between passive smoke exposure and urinary estrone conjugate levels among premenopausal women (Chen et al. 2005). Active smoking, but not passive smoking, has been associated with decreased antimullerian hormone levels, a marker of ovarian function (Plante et al. 2010). In another study, follicle-stimulating hormone levels, also a marker of ovarian function, were higher among active smokers (66% higher) and passive smokers (39% higher) relative to women with no active or passive smoking exposure (Cooper et al. 1995). Several investigations report negative associations between age at menopause and active smoking, but not passive smoking (Cooper et al. 1999, Cramer et al. 1995, Mikkelsen et al. 2007), though one earlier study did report a negative association between age at menopause and passive smoking (Everson et al. 1986). Active smoking, but not passive smoking, was associated with reduced mammographic density in a 2010 investigation (Butler et al. 2010).

Assessment of Smoking Patterns in Epidemiological Investigations

In epidemiological studies, information on smoking habits is ascertained through self-
report and classified in a variety of ways (Terry and Rohan 2002). The most crude smoking exposure categorizations are ever/never smoking or current/former/never smoking. Some studies, especially those conducted more recently, have collected more comprehensive information on smoking patterns, including estimates of frequency, intensity, duration and recency of smoking (Terry and Rohan 2002).

Reporting of smoking patterns is subject to recall bias and social desirability bias (Brigham et al. 2010). Self-reported smoking exposure has been validated against measured cotinine levels (Nasca and Pastides 2008, Patrick et al. 1994). However, cotinine reflects only recent smoking, and cannot differentiate between levels of tobacco smoke exposure (Nasca and Pastides 2008, Patrick et al. 1994). Research has demonstrated long-term consistency of smoking reports that were collected at different time points across the life course, though concordance is not absolute (Brigham et al. 2010).

**Associations with Breast Cancer**

An association between cigarette smoke and breast cancer is biologically plausible based on knowledge of relevant mechanisms and pathways, cigarette smoke components, and toxicokinetics (Hecht et al. 2002). Furthermore, smoking is a risk factor for cancers in other organs which are not directly exposed to tobacco smoke, such as the bladder and pancreas (Brennan et al. 2000, Lynch et al. 2009, Terry and Rohan 2002). The association between smoking and breast cancer has been widely studied (Terry and Rohan 2002). However, results are inconsistent across investigations (positive, null, and negative) (Palmer and Rosenberg 1993, Terry and Rohan 2002). A review by Terry and Rohan (2002) concluded that a true negative association between smoking and breast cancer is extremely unlikely.

More consistently positive associations are found among women with certain genetic

Women may be more susceptible to initiating breast carcinogens prior to pregnancy because breast epithelial cells have not undergone terminal differentiation and are at "a peak of cell replication" (Nasca and Pastides 2008, Russo and Russo 2004). Differing results for passive versus active smoking may be due to (1) a potential antiestrogenic effect of smoking that could be present upon active but not passive exposure levels, as described above (Band et al. 2002), (2) differences in duration or timing of exposure to passive versus active smoke, as women were historically more likely to be exposed earlier in life to passive than to active smoke, (3) the failure of many studies of active smoking to remove passive smokers from the referent group, which could attenuate results (Morabia et al. 1996, Terry and Rohan 2002), and (4) competing causes of death potentially resulting from active but not passive smoking exposure levels (Dr. Marilie Gammon, personal communication 2010). It should be noted that "the California Environmental Protection Agency has concluded that regular exposure to secondhand smoke is causally related to breast cancer diagnosed in younger, primarily premenopausal women" (ACS 2009-2010) (Miller et al. 2007).

Cigarette smoking was associated with the presence of breast tumor \textit{p53} mutations in one study (Conway et al. 2002), but this result was not confirmed in a larger investigation (Mordukhovich et al. 2009). A 2004 literature review concluded that cigarette smoking is not
differentially associated with breast tumor hormone receptor status subtypes (Althuis et al. 2004), although individual investigations have reported evidence of differential associations (e.g. Britton et al. 2002, Gammon et al. 2004b).

**Grilled and Smoked Meat**

**PAH Content of Grilled and Smoked Meat**

PAHs are found on or near the surface of grilled, barbecued and smoked meat (Morris and Seifter 1992 Steck et al. 2007). Smoke is formed during the incomplete combustion of hydrogen and carbon in fat dripping onto a heat source; the smoke rises and deposits PAHs on meat (IARC 2010, Kazerouni et al. 2001, Lijinsky 1991, Steck et al. 2007). In addition, PAHs are deposited on meat during preservation by smoke curing (IARC 2010). PAHs are formed directly on charred meat (Jagerstad and Skog 2005). Cooked meat also contains non-PAH carcinogens, such as heterocyclic amines (Taylor et al. 2009).

PAH levels in cooked meat depend on several factors, including cooking method, temperature and duration, doneness level, distance from the heat source, the amount of fat on the meat, and whether or how much fat dripped onto the heat source (IARC 2010, Kazerouni et al. 2001, Lijinsky 1991, Lijinsky and Shubik 1965, Morris and Seifter 1992, Ramesh et al. 2004). PAH levels in smoked meat depend on the specific smoking technique used (Gomaa et al. 1993, IARC 2010). It is estimated that, on average, an individual's dietary intake of PAHs in the US is less than 2 μg per kg of food (Agency for Toxic Substances and Disease Registry [ATSDR] 1995). The highest levels of dietary PAHs are found in charred meat (up to 10-20 μg per kg of food) (Castaño-Vinyals et al. 2004, Phillips 1999).

**Assessment of Grilled and Smoked Meat Intake in Epidemiological Studies**

In epidemiological studies, dietary habits are generally assessed through food frequency
questionnaires or other dietary questionnaires. To a more limited extent, nutritional epidemiology studies may use biomarkers of exposure including DNA adducts and urinary metabolites to evaluate certain types of short-term dietary intake (IARC 2010, Ramesh et al. 2004, Roth et al. 2001, Strickland et al. 2002).

Study questionnaires can be used to reconstruct dietary PAH exposures through questions on intake patterns and cooking methods. Researchers can evaluate associations between health outcomes and specific PAH-containing foods (such as grilled and smoked meat), or can use a combination of questionnaire responses and databases of measured PAH concentrations in foods to construct a dietary PAH exposure index (Kazerouni et al. 2001, Sinha et al. 2005, Steck et al. 2007).

Regarding the latter approach, while PAH content in foods can be physically measured by assessing concentrations of several PAHs (Guillen 1994, Phillips 1999), this is difficult in practice given that PAH profiles differ between foods and not all PAHs are easily measured (IARC 2010). Another option is to assess benzo[a]pyrene as a surrogate for all dietary PAHs (Kazerouni et al. 2001) (IARC 2010). Benzo[a]pyrene is both a strong carcinogen and a reasonable surrogate for overall dietary PAHs because of its presence in many food items (IARC 2010). A 2001 investigation reported a correlation of $r = 0.87$ ($p = 0.0001$) between dietary benzo[a]pyrene concentrations and the combined levels of 14 other PAHs in a variety of foods. The correlation was even greater ($r = 0.98; p = 0.0001$) for PAHs that are carcinogenic in animals (Kazerouni et al. 2001) (IARC 2010). Most studies of meat and breast cancer have evaluated associations with foods rather than an exposure index for specific chemicals. The Long Island Breast Cancer Study Project evaluated associations with lifetime intake of grilled and smoked meat and also constructed a benzo[a]pyrene dietary index (Steck et al. 2007).
Dietary questionnaires are hampered by a number of methodological limitations, including recall bias (Byers et al. 1983). Recall bias may be especially problematic when assessing dietary intake in the distant past. A number of studies have attempted to validate long-term dietary recall and reported correlation coefficients ranging from 0.13 to 0.59 (Bakkum et al. 1988, Byers et al. 1983, Jensen et al. 1984, Lindsted and Kuzma 1989, Maruti et al. 2005, Sobell et al. 1989, Wu et al. 1988) (Steck et al. 2007). The resultant misclassification can bias results towards or away from the null for a multi-level dietary variable (Steck et al. 2007). Previous research suggests that recall bias for dietary intake is generally not differential by cancer case-control status (Byers et al. 1983, Freidenreich et al. 1992, Holmberg et al. 1996, Jensen et al. 1984, Steck et al. 2007). However, one breast cancer study found that recall bias patterns for reporting of meat intake differed by case-control status, though not in a consistent direction with respect to different meat variables (Holmberg et al. 1996).

In most epidemiological studies of meat consumption and breast cancer risk, questionnaires focused only on recent meat intake, most commonly in the past year (Steck et al. 2007). Longer-term exposure may be more relevant to breast carcinogenesis (Clark et al. 1997). In addition, many dietary questionnaires do not (or do not adequately) evaluate cooking methods (Steck et al. 2007), which does not allow differentiation between overall meat intake and intake of grilled, smoked or barbecued meat specifically (Steck et al. 2007). Studies that do evaluate cooking techniques or doneness preferences differ in the method of assessing this information (Sinha et al. 2000, Steck et al. 2007, Zheng et al. 1998). Resulting variation in reporting dietary habits may impact effect estimates.

**Associations with Breast Cancer**

Studies evaluating associations between meat consumption and breast cancer report

Most (Dai et al. 2002, Deitz et al. 2000, Krajinovic et al. 2001, Zheng et al. 1998, 1999, 2001, 2002), but not all (Delfino et al. 2000, Gertig et al. 1999), epidemiological studies do report positive associations between breast cancer and intake of well-done red meat both in general and among women with certain genetic variants. The two studies reporting null results for well-done red meat consumption also found null or negative associations between breast cancer and grilled, barbecued, or charred red meat intake over the course of one year (Delfino et al. 2000, Gertig et al. 1999). Few women reported eating red meat in the study by Delfino and colleagues (2000). Another investigation found null results for total (red and white) well-done or charred meat consumption (Holmes et al. 2003). A recent study from the Long Island Breast Cancer Study Project evaluating lifetime intake of grilled and smoked meat found positive associations with postmenopausal breast cancer, which were strongest among women with low fruit and vegetable intake (Steck et al. 2007). Elevated effect estimates were reported for intake of red, but not white, grilled and smoked meat (Steck et al. 2007).

**PAH-DNA Adducts**

**Biomarkers of PAH Exposure**

Biomarkers of exposure or dose are measurements of chemicals or their metabolites in
the body (e.g. in tissues, blood, or urine) (Godschalk et al. 2003). PAH biomarkers include blood-protein adducts (most commonly assessed in albumin or hemoglobin), urinary PAH metabolites (primarily 1-hydroxypyrene), and DNA adducts in blood or tissues (Bostrom et al. 2002, Castaño-Vinyals et al. 2004).

A 2003 review by Godschalk and colleagues concluded that, for PAHs, "the most promising biomarker seems to be the measurement of DNA adducts, since it takes into account individual differences in exposure, absorption and distribution of the chemical, its metabolism into DNA reactive forms, detoxification [of] reactive intermediates, as well as cell turnover and repair of DNA damage" (Godschalk et al. 2003). Since PAH-DNA adducts reflect DNA damage levels, they may be considered an indicator of effective biological PAH dose, rather than purely of exposure (Beyea et al. 2006, Binkova et al. 1995, Nesnow et al. 1993). PAH-DNA adducts have been found in human breast tissue and human breast milk (Bostrom et al. 2002, Gammon and Santella 2008, Moore et al. 1987, Ramesh et al. 2004, Rundle et al. 2000a, Santella 1999, Straif et al. 2005, Terry and Rohan 2002).

Due to high cell turnover rates (usually a few months to a few years), PAH-DNA adducts generally represent recent exposures in both target and surrogate (e.g. blood) cells (Gammon et al. 2002c, Shields et al. 1992). PAHs are lipophilic and can thus also "accumulate in adipose tissue to be released over time" (Rundle et al. 2000a). Hence, PAH adducts may potentially reflect exposures from the more distant past (Obana et al. 1981, Rundle et al. 2000a). Nevertheless, no known biomarkers represent long-term, historical PAH exposure or dose (Gammon et al. 2002c).

**PAH Sources as Predictors of Adduct Levels**

Many studies report associations between PAH-related DNA adducts and ambient

Ambient environmental PAHs can affect PAH-DNA adduct levels more strongly than dietary sources, active smoking, or occupational exposures (Beyea et al. 2006, Eder 1999), despite the fact that the average person’s exposure to PAHs from air pollution is lower than from these other sources (Castaño-Vinyals et al. 2004). It is unclear why this is the case. However, this fact does illustrate that exposure level is not the only consideration with respect to biological relevance. Other factors including exposure route, actual internal dose, differences in metabolism or DNA repair, and other components of the PAH mixture may be of vital importance with respect to disease causation and other biologic effects (Bostrom et al. 2002).

Not all studies have found associations between PAH sources and adduct levels (Gammon et al. 2004a, Georgiadis et al. 2001, Perera et al. 1995, Scherer et al. 2000, Shantakumar et al. 2005). This may be due to differing biological relevance of various PAH exposure sources, differences in PAH exposure levels between investigations, methodological limitations, small sample sizes, differing study population characteristics relevant to PAH metabolism or DNA repair, or the influence of unmeasured PAH exposures (Bostrom et al. 2002, Eder 1999, Shantakumar et al. 2005). It has been hypothesized that inconsistent associations between PAHs and adducts may be an indication that adduct formation is more a function of individual susceptibility than exposure level (Gammon et al. 2004a). Despite correlations
between PAH-related exposures and PAH-DNA adducts in many studies, individuals' adduct levels are known to vary significantly at similar exposure levels. This “suggests that individuals respond differently as a result of genetic susceptibility related to activation or detoxification of carcinogens, DNA repair capacity, and other lifestyle or dietary factors” (Santella 1999).

A 2004 review that pooled results from twelve studies found a statistically significant correlation between ambient environmental benzo[a]pyrene levels and mean PAH-DNA adduct levels in blood cells (Castaño-Vinyals et al. 2004). This review concluded that "PAH-DNA adducts can be usefully applied to assess environmental exposure to PAHs at a group level" (Castaño-Vinyals et al. 2004), and that biomarker measurement can discriminate between small differences (as low as 5 ng/m³) of monitored personal exposure to ambient benzo[a]pyrene (Castaño-Vinyals et al. 2004). In this pooled analysis, the correlation coefficient varied with respect to adduct measurement strategy. Specifically, the correlation between ambient environmental benzo[a]pyrene and PAH-DNA adducts was stronger when using enzyme-linked immunosorbent assay (ELISA) (r = 0.99; P = 0.02) than when using ³²P-postlabeling (r = 0.406; P = 0.21, overall combined correlation: r = 0.60; P = 0.04) (Castaño-Vinyals et al. 2004). It should be noted that ambient benzo[a]pyrene levels happened to be higher in studies using ELISA than in studies using ³²P-postlabeling (Castaño-Vinyals et al. 2004).

**Laboratory Assessment of PAH-DNA Adducts**

PAH-DNA adducts can be detected in circulating blood cells and in biopsied healthy or malignant target tissue, including breast tissue (Bostrom et al. 2002). PAH-DNA adducts are usually assessed in the blood, which is representative of "the systemic concentration of the parent compound and/or reactive metabolites" (Bostrom et al. 2002). This may potentially attenuate associations between adduct levels and cancer risk relative to measurement in the target tissue.
Researchers report correlations between PAH-DNA adduct levels in the blood and at least one group of target tissues - the lung and larynx (Bostrom et al. 2002, Szyfter et al. 1994, Wiencke et al. 1995). Other studies of correlations between blood and tissue adducts in the lung and skin have shown null or mixed results (Godschalk et al. 1998a,b, Poirier et al. 2000, Santella et al. 1999, Spivack et al. 1997, van Schooten et al. 1997). This association has not been evaluated for breast tissue.

A number of laboratory methods are available to assess PAH-DNA adduct levels, including $^{32}$P postlabeling of modified nucleotides, mass spectrometry, fluorescence spectroscopy, immunohistochemistry, and competitive ELISA (Bostrom et al. 2002, Gammon et al. 2000c, Poirier et al. 2000). Several other methods of adduct detection exist, but are used rarely (never for breast cancer research) (Santella et al. 1999). These methods are not discussed further in this document.

ELISA measures specific PAH-DNA adducts by using targeted antibodies (Bostrom et al. 2002) including against "benzo[a]pyrene and structurally related PAH diol epoxide-DNA adducts" (Gammon et al. 2002c). Like ELISA, fluorescence spectroscopy and mass spectrometry also identify and quantify structurally specific DNA adducts (Bostrom et al. 2002, Casale et al. 2001, Rojas et al. 1994, 1998). In contrast, $^{32}$P-postlabeling measures general bulky aromatic or hydrophobic DNA adducts (Castaño-Vinyals et al. 2004, Phillips and Arlt 2007). Thus, this method is not specific to PAH-DNA adducts. For example, it also measures adducts from heterocyclic amines (Castaño-Vinyals et al. 2004, Poirier and Weston 1996). However, postlabeling is a more sensitive method than ELISA, fluorescence spectroscopy, or mass spectrometry (Bostrom et al. 2002, IARC 2010).

Immunohistochemistry is another option, allowing for the detection of adducts in target
tissue, such as breast tissue. Immunohistochemical assays are inexpensive and have good sensitivity and specificity for PAH-DNA adducts (IARC 2010, Romano et al. 1999, Rundle et al. 2000a, Santella 1999). Both ELISA and immunohistochemistry require “the production of antibodies, prior knowledge of the adduct being measured and relatively large amounts of DNA for sensitivity” (IARC 2010) (Santella et al. 1999). ELISA is generally more sensitive than immunohistochemistry (Santella et al. 1999). Of all available methods, ELISA may be “most suited to large-scale epidemiology studies or clinical evaluations because it is a rapid and low-cost procedure not requiring radiolabeled compounds” (Casale et al. 2001).

PAH-DNA adduct assays do not measure all forms of PAH-induced DNA damage (Casale et al. 2001). For example, the Long Island Breast Cancer Study Project measured PAH-DNA adducts using an antibody against PAH diol epoxide adducts forming at the N2 position of guanine (Gammon et al. 2002c, Mordukhovich et al. 2009). This assay does not recognize PAH quinone-DNA adducts, abasic sites left by depurinating adducts, oxidative lesions, or diol epoxide adducts at other positions. Assays which detect all bulky adducts will still fail to detect abasic sites and oxidative lesions.

Population-based studies of PAH-DNA adducts and breast cancer assessed adducts in circulating mononuclear cells using ELISA (Gammon et al. 2002c, 2004a) or in leukocytes using $^{32}$P-postlabeling (Saieva et al. 2011). The Long Island Breast Cancer Study Project chose to use ELISA because of its specificity to PAH-DNA adducts and the speed and convenience of the assay, all assets for a large investigation (Gammon et al. 2002c). All other studies of PAH-related DNA adducts and breast cancer were hospital-based and measured adducts directly in breast tissue (tumor and/or adjacent tissue for cases), using control participants who underwent either reduction mammoplasties (Li et al. 1996, 2002, Perera et al. 1995) or surgery for certain

**Associations with Breast Cancer**

PAH-related DNA adducts are consistently related to breast cancer both in small hospital-based case-control investigations (Li et al. 1996, 2002, Perera et al. 1995, Rundle et al. 2000a) and in larger population-based case-control studies (Gammon et al. 2002c, 2004a). For example, Rundle and colleagues (2000) found that breast tumor tissue was more than twice as likely to contain elevated PAH-DNA adduct levels as breast tissue from women with benign breast disease (odds ratio [OR] = 2.56; 95% confidence interval [CI]: 1.05-6.24). Results were of borderline significance for adducts in adjacent, non-tumor tissue (OR = 1.97, 95% CI: 0.94 - 4.17) (Rundle et al. 2000a). Similar or greater effects were observed in the other hospital-based studies (Li et al. 1996, 2002, Perera et al. 1995).

A population-based case-control study (the Long Island Breast Cancer Study Project) reported more modest positive associations between breast cancer and PAH-DNA adducts in circulating mononuclear cells (OR = 1.41, 95% CI: 1.07, 1.86 for the highest vs. lowest quantile, OR = 1.29, 95% CI: 1.05, 1.58 for detectable vs. nondetectable adduct levels) (Gammon et al. 2004a). This reduction in effect estimates relative to hospital-based investigations is likely related to increased sample size and therefore improved stability of results (Gammon et al. 2004a). The fact that the hospital-based studies assessed DNA adducts directly in breast tissue rather than in blood cells may also be relevant (Perera and Rundle 2003). No dose-response relationship was observed when evaluating quantiles of PAH-DNA adduct levels in relation to breast cancer risk (Gammon et al. 2004a). This may be due to (1) a threshold effect of PAH-
DNA adducts on breast cancer risk, (2) greater importance of individual susceptibility to PAHs than of PAH exposure level, or (3) a potentially non-linear dose-response relationship between adducts and breast cancer (Dr. Jan Beyea, personal communication 2009, Gammon et al. 2004a). A recent Italian nested case-control study (292 cases, 292 controls) failed to find an association between bulky adducts in leukocytes and breast cancer risk (Saieva et al. 2011).

TRAFFIC AND AIR POLLUTION

Sources of Exposure to Ambient Environmental PAH

Ambient PAHs, both indoor and outdoor, are emitted from anthropogenic sources including vehicular traffic, industrial processes, power generation, heating, incinerators, wood and coal stoves, cooking, candles, incense, tobacco smoke, space heaters, and leaf and garbage burning (Besaratinia et al. 2002, Boffetta 1997, Bostrom et al. 2002, Friedman and Calabrese 1977, IARC 2010, Lioy and Greenberg 1990, Mumtaz et al. 1996, Narvaez et al. 2008, Ramesh et al. 2004, Wallace 2000). In addition to trucks and cars, mobile sources of PAH emissions include "aircraft, shipping, railways...off-road vehicles, and machinery" (Ravindra et al. 2008). Natural PAH sources such as forest fires and volcanoes make a small contribution to ambient PAH levels at ground level, which is heavily outweighed by anthropogenic emissions (Bostrom et al. 2002). Because of the importance of anthropogenic PAH emissions relative to PAHs from natural sources, PAH levels in water sediments are strongly correlated with the historical onset of fossil fuel use (Bostrom et al. 2002).

Measured PAH air concentrations (Gigliotti et al. 2005, Morris and Seifter 1992) and urinary PAH metabolite levels (Adonis et al. 2003, Kanoh et al. 1993, Kuo et al. 2004) are higher in urban areas than in more rural locations. Vehicular traffic emissions are often the primary source of ambient PAH in urban and suburban areas (Bostrom et al. 2002, Dubowsky et al. 1999,
Dunbar et al. 2001, IARC 1989), though wood smoke is the largest source of ambient PAHs in the US overall (Bostrom et al. 2002). PAHs emitted from wood burning show relatively little carcinogenic activity, however, in contrast to PAHs in vehicular traffic emissions (Boffetta et al. 1997, Lewtas 1993, Lewtas et al. 1992, Lewtas and Gallagher 1990). Wood burning is also more likely to occur away from areas of high population density (Shen et al. 2011). PAHs, primarily bound to course particles (> 10 μm in diameter) (Bostrom et al. 2002), are also found in road dust originating from brake linings, tires, and asphalt (Bostrom et al. 2002, Ravindra et al. 2008).


**Ambient PAH Levels and Properties**

Ambient PAH levels vary significantly across regions (IARC 2010). Concentrations tend to be high in large cities (IARC 2010). Measured ambient benzo[a]pyrene levels range from 0.01 to 100 ng/m$^3$, and measured overall ambient PAH levels range from less than 5 ng/m$^3$ to hundreds of ng/m$^3$ (Castaño-Vinyals et al. 2004, Georgiadis and Kyrtopoulos 1999; IARC 2010, Vyskocil et al. 1997). A 2010 review reports typical levels of a few ng/m$^3$ for ambient benzo[a]pyrene and a typical range of 1 to 30 ng/m$^3$ for total ambient PAHs, excluding naphthalene (IARC 2010). PAH air concentrations are substantially higher in urban than in more rural areas (Gigliotti et al. 2005, Morris and Seifter 1992).

Most studies report that ambient PAH levels are highest in the winter (Flessel et al. 1991, Harrison et al. 2009, Moller et al. 1996, Srogi 2007, Topinka et al. 2000). This is likely due to lower temperatures and solar radiation levels, leading to reduced PAH degradation. Other potential contributing factors vary by location and include summer rainfall facilitating PAH

Ambient PAH levels have decreased significantly over the past several decades due to the widespread introduction of catalytic converters in gasoline vehicles (Bostrom et al. 2002). The lowered PAH emissions per vehicle were in part counteracted by increased vehicle use during the same time period (Beyea et al. 2008). For example, a 2008 report described an approximately "15-fold decline in benzo[a]pyrene and other PAH cruise emission factors for the US vehicle fleet following introduction of automotive pollution controls," whereas "overall vehicle miles traveled in the US increased by about a factor of 2.7 between 1970 and 2005" (Beyea et al. 2008) (Davis and Diegel 2007). Despite the historical decline in US traffic PAH emissions, it is important to evaluate associations between traffic PAHs and breast cancer for the following reasons: (1) many US women are still exposed to high levels of traffic PAHs, such as women living near highways, bus depots, and busy intersections, and (2) overall vehicular PAH emission levels remain high in other countries.

PAHs can be found in both gaseous and particulate phases, and are therefore considered semivolatile organic compounds (Zielinska et al. 2004). Larger PAH molecules, such as benzo[a]pyrene, are usually particle-associated rather than in the gas phase and are generally much more carcinogenic than lighter-weight PAHs (Beyea et al. 2006, Bostrom et al. 2002, IARC 2010, Zielinska et al. 2004). Airborne PAHs from gasoline vehicle emissions are primarily bound to particles (Bostrom et al. 2002, Westerholm and Egebäck 1994). The relative proportion of particle and gas-phase PAHs is affected by meteorologic conditions (i.e. relative
humidity and temperature), the properties of specific PAHs in ambient mixtures, and the
Ambient gas-phase PAHs “generally have durations of less than a day, whereas particle-
associated PAHs may persist for weeks and undergo long-range atmospheric transport” (IARC

It is reported that 90-95% of particle-associated PAHs are bound to particles less than 3.3
μm in diameter (Bonner et al. 2005, Ravindra et al. 2001). A 1998 study reported that almost all
PAHs from gasoline vehicle emissions are bound to ultrafine particles (< 0.12 μm in diameter)
(Bostrom et al. 2002, Miguel et al. 1998). Relative to larger particles, small particles are able to
penetrate more deeply into the respiratory tract, are absorbed into the circulation more quickly,
and induce systemic oxidative stress and inflammation more efficiently (Li et al. 2003, Miller
1973).

*Atmospheric Fate of Ambient PAH*

Airborne PAHs are oxidized and degraded in the atmosphere through photochemical and
chemical transformation (i.e. due to interactions with radiation and other air pollutants), and are
removed from the atmosphere via dry and wet deposition (Bostrom et al. 2002). Deposition
processes are influenced by particle size and relative concentrations of particle-associated and
gaseous PAHs (Bostrom et al. 2002). The efficiency of chemical transformation depends on the
concentrations of other air pollutants, the relative proportion of particulate and gas-phase PAHs,
and on meteorological conditions (Bostrom et al. 2002). Photochemical transformation tends to
decrease in the winter relative to other seasons because of lower temperatures and solar radiation
levels (Bostrom et al. 2002).

Environmental monitoring and measurement methods are not well developed for PAHs (Bostrom et al. 2002). Monitoring of specific PAHs “often requires extensive separation schemes because of their lack of distinct functional groups, the existence of numerous structural isomers and the need to analyze PAHs in diverse environmental matrices” (IARC 2010). Unlike US EPA criteria pollutants such as carbon monoxide or ozone, environmental ambient PAHs are monitored relatively infrequently and ambient PAH levels are thus not well quantified (Bostrom et al. 2002).

In epidemiologic research, ambient PAH exposure assessment methods include the use of personal monitors, measurement of PAH biomarkers of exposure or dose, and ascertainment of PAH-related exposures, such as traffic exposure variables for research in the general population, or coal-tar pitch volatiles in occupational studies, which may be used as surrogates of ambient PAH exposure level (Gunier et al. 2006, IARC 2010). However, personal monitors are expensive and present other logistical difficulties, PAH biomarkers integrate non-ambient PAH sources and are likely to be influenced by individual-level factors and characteristics in addition to exposure levels (Shantakumar et al. 2005; see Chapter I), and surrogates of ambient PAH concentrations, such as distance to the nearest major road, are often crude.
More than 100 PAHs have been isolated from air. Studies that measure ambient PAHs directly usually report concentrations for a small number of PAHs, thereby potentially excluding PAHs relevant to health effects of interest (Bostrom et al. 2002, Castaño-Vinyals et al. 2004, IARC 2010). Furthermore, the collection of measured PAHs differs across investigations (Bostrom et al. 2002, Castaño-Vinyals et al. 2004). This is in large part due to the difficulty of finding an index PAH, since PAH profile varies with respect to both the specific exposure source and the temperature of combustion (Bostrom et al. 2002, Castaño-Vinyals et al. 2004, IARC 2010).

Most often, only benzo[a]pyrene is measured as a surrogate for all ambient PAHs (Bostrom et al. 2002, Castaño-Vinyals et al. 2004). On the one hand, "the presence in the environment of other PAHs, some of which are carcinogenic, does not allow an accurate estimation of the risk linked to a PAH mixture on the basis of [benzo[a]pyrene] concentration alone" (Castaño-Vinyals et al. 2004). Similarly, in a real-life setting, “most if not all exposures...to PAHs involve complex organic mixtures” and “metabolism and bioactivation of individual PAH compounds are influenced by the presence of other PAHs” (Ramesh et al. 2004) (Goldstein 2001, Warshawsky 1999). Traffic emissions include significant concentrations of benzo[g,h,i]perylene, pyrene, fluoranthene, phenanthrene and other PAHs (Bostrom et al. 2002). Monitoring only benzo[a]pyrene inevitably leads to some exposure misclassification (Bostrom et al. 2002, Castaño-Vinyals et al. 2004).

Nevertheless, benzo[a]pyrene is one of the most carcinogenic PAHs and is common to all ambient PAH sources (Bostrom et al. 2002, Castaño-Vinyals et al. 2004). Benzo[a]pyrene is representative of ambient PAH exposure from vehicular exhaust (Beyea et al. 2008, Bostrom et al. 2002, Fertmann et al. 2002), and is a good surrogate for PAH carcinogenicity (Bostrom et al. 2002, Castaño-Vinyals et al. 2004).
Furthermore, the Long Island Breast Cancer Study Project reports that soil benzo[a]pyrene levels are strongly correlated with other high molecular weight soil PAHs, which are generally more carcinogenic than lower weight PAH molecules (Beyea et al. 2006, Bostrom et al. 2002). For example, Pearson correlation coefficients were greater than 0.9 for benzo[b]fluoranthene, benzo[ghi]perylene, benzo[k]fluoranthene, dibenz[a,h]anthracene and indeno[1,2,3-cd]pyrene, and were greater than 0.8 for acenaphthylene, anthracene, benz[a]anthracene, chrysene and fluorine (Beyea et al. 2006). With respect to historical reconstruction of traffic emissions exposure, benzo[a]pyrene is the only PAH for which historical data is consistently available (Beyea et al. 2008). A 2008 report indicated that the pattern of historical decline of ambient PAH levels was similar between benzo[a]pyrene and other traffic-related PAHs, including benzo[e]pyrene, benzo[ghi]perylene, chrysene, coronene, and perylene (Beyea et al. 2008).

Cancer risk assessment with respect to ambient PAH exposure is difficult given the fact that PAHs are rarely monitored, are almost always components of complex mixtures including varied carcinogenic PAHs and non-PAH carcinogens, and are mainly bound to particles, which can exert carcinogenic effects independent of PAH content (Bostrom et al. 2002, Valavanidis et al. 2008). Risk assessment for PAHs with respect to various disease endpoints is largely based on animal studies in which one surrogate PAH is examined individually on a short-term basis, and from studies of high levels of complex occupational exposures, such as coal-tar pitch volatiles (ATSDR 1995).

Several research groups and organizations, including the World Health Organization (WHO) and the California EPA, have calculated lifetime risk estimates for an increase in lung
cancer risk in relation to a rise in ambient benzo[a]pyrene levels (Bostrom et al. 2002, CA EPA Air Resources Board [CARB 1994], WHO 2000). For example, the WHO estimates "the lung cancer risk from a lifetime exposure to PAHs in ambient air at \(8.7 \times 10^{-5}\) per ng/m\(^3\) benzo[a]pyrene" (Bostrom et al. 2002) (WHO 2000). Similar risk slopes have not been estimated for breast or other cancers.

**Regulations for Ambient PAH and Related Pollutants**

The US EPA does not regulate ambient environmental PAH levels, though it does set standards for several PAHs in drinking water (US EPA 2009). The US Occupational Health and Safety Administration (OSHA) limits occupational exposure to coal tar pitch volatiles, a PAH-heavy mixture (Han and Naeher 2006). Some other countries also limit occupational exposures to coal-tar pitch volatiles or particulate PAHs, but none limit ambient environmental PAH levels (IARC 2010).

The US EPA regulates levels of ambient particulate matter (PM; US EPA 2011b), which often contains adsorbed PAH molecules (conversely, ambient heavy mass PAHs are mostly physically associated with particles) (Bostrom et al. 2002). The EPA also regulates ambient levels of the traffic-associated pollutants carbon monoxide, ozone, nitrogen oxides and sulfur dioxide (US EPA 2011b), as well as tailpipe exhaust emissions of PM, carbon monoxide, nitrogen oxides, non-methane organic gases, and volatile organic compounds (VOCs) (Environment and Human Health, Inc. [EHHI] 2006, US EPA 1994a). It should be noted that, in practice, ambient pollutant concentrations often exceed regulatory standards (EHHI 2006). Also, air quality standards may need to be updated given evidence of health effects of traffic-related air pollutants even at levels that fall below federal standards (EHHI 2006).

Traffic exposure is widespread (Han and Naeher 2006) and is usually the main source of air pollution in urban areas (Maitre et al. 2002). Traffic emissions vary in persistence, "physical and chemical properties, composition, and toxicity of fuel mixtures" across geographic and demographic regions (Boothe and Shendell 2008) (Verma and des Tombe 2002). This is due to a number of factors including average vehicle age, type, and maintenance status, fuel composition, presence of emissions control technologies such as catalytic converters, traffic and driving patterns, governmental regulations, and roadway conditions, "designs, grades, and distributions" (Boothe and Shendell 2008) (IARC 1989).

Particles from gasoline and diesel vehicle emissions differ with respect to size and
surface properties (IARC 1989). Organic compound emissions from gasoline or diesel sources differ mainly in quantity. For example, diesel engines emit greater levels of PM and nitroarenes relative to gasoline vehicles with catalytic converters, given comparable power output (2-40 times and 20-30 times higher, respectively) (IARC 1989). For PAHs in particular, emission levels are similar between diesel vehicles and gasoline vehicles without catalytic converters given similar power output, but are greatly reduced (by more than 10 times) in gasoline vehicles with catalytic converters (IARC 1989).

Measured ambient vehicular exhaust levels are highest near roads, with concentrations dropping to background levels within approximately 150-300 meters (Boothe and Shendell 2008) (Gilbert et al. 2005, Zhu et al. 2002, 2006). This decline is due to "evaporation of volatile constituents, atmospheric dispersion, and coagulation" (Boothe and Shendell 2008) (Zhu et al. 2002).

**Influences on Traffic PAH Emissions**

Traffic PAH emission levels differ based on driving conditions and traffic patterns (Boothe and Shendell 2008, IARC 1989). Vehicular PAH emissions are related to driving speed, with greater emissions at low and high speeds (Begeman and Colucci 1970, Kotin et al. 1954, Westerholm et al. 1992), and are higher during transient rather than steady-state driving cycles (Kado et al. 2005) (Shen et al. 2011). Emissions are elevated at traffic intersections due to increased acceleration and deceleration at these sites (Beyea et al. 2006, 2008, Sculley 1989, Sheu et al. 1996a,b), and are also greater under cold-engine rather than warm-engine conditions, discussed in more detail below (Beyea et al. 2006, 2008, Bostrom et al. 2002, Ravindra et al. 2008, Shen et al. 2011).

Emission levels also differ based on fuel characteristics and vehicle properties. PAH
emissions differ with respect to fuel type (diesel versus gasoline, as described below) (Bostrom et al. 2002, Shen et al. 2011). Also, both diesel and gasoline vehicle PAH emission levels are positively associated with increased concentrations of PAHs in the respective fuels (Bostrom et al. 2002, Shen et al. 2011, Ravindra et al. 2008). Vehicular PAH emission levels are decreased greatly by the presence of a catalytic converter (Beyea et al. 2006, 2008, Bostrom et al. 2002, Shen et al. 2011). Emissions are associated with engine condition and maintenance as well increased mileage (Ravindra et al. 2008). PAH emission levels also vary with respect to "vehicle age and load, lubricant oil (Ravindra et al. 2008)" (Shen et al. 2011) and outdoor ambient temperatures (Bostrom et al. 2002, Shen et al. 2011). Information regarding variation in PAH emission levels is derived from tunnel studies, road-side measurements, and dynamometer test beds (Beyea et al. 2008, Ravindra et al. 2008, Shen et al. 2011).

Both gaseous and particulate PAHs are found in vehicle exhaust, but particulate PAHs are much more prevalent in gasoline vehicle emissions (Bostrom et al. 2002, Han and Naeher 2006, Miguel and Friedlander 1978, Morris and Seifter 1992). Gasoline-powered cars tend to emit higher molecular weight PAHs relative to heavy-duty diesel vehicles (e.g. coronene and benzo[a]pyrene vs. fluoranthene and pyrene, respectively) (Bostrom et al. 2002). Heavy-duty diesel vehicles generally emit greater amounts of PAH per vehicle for the same distance traveled relative to gasoline vehicles without catalytic converters (Boffetta et al. 1997, Han and Naeher 2006), though gasoline-fueled passenger vehicles “without catalytic converters and diesel engines of similar power output produce similar quantities of polycyclic aromatic hydrocarbons per kilometer" (IARC 1989). Gasoline cars with catalytic converters emit fewer PAHs than any other vehicle type (Boffetta et al. 1997, IARC 1989), since catalytic converters reduce vehicular PAH emissions by greater than 10 times (IARC 1989). Currently, it is likely that diesel vehicles
and older gasoline vehicles without catalytic converters emit the majority of traffic-related PAHs (Bostrom et al. 2002). Studies indicate that diesel engine PAH emissions are mainly comprised of less carcinogenic lighter weight PAHs, and that gasoline engines have historically been the major vehicular source of heavy mass PAHs such as benzo[a]pyrene in the US (Begeman and Colucci 1970, Beyea et al. 2008, Bostrom et al. 2002, Miguel et al. 1998).

Although cold-engine emissions occur for only a short time and distance after starting a vehicle, they may nevertheless account for more than 50% of PAH emissions from gasoline cars (Bostrom et al. 2002). Relative contributions from cold-engine emissions differ with respect to time and location (Beyea et al. 2006). For example, the contribution of cold-engine emissions is higher in residential areas relative to highways, and at times when people are most likely to be starting their cars, such as prior to their morning commute (Beyea et al. 2006). The distribution of gas and particle-phase PAHs does not appear to differ based on engine temperature conditions (Ravindra et al. 2008).

**Traffic and Air Pollution Exposure Assessment**

Most epidemiological studies assess air pollution exposure through either sparse pollutant monitoring data or through indirect measures such as distance to a pollution source (Beyea 1999, Boothe and Shendell 2008, Heinrich et al. 2005, Savitz and Feingold 1989). Environmental ambient PAH concentrations are not commonly monitored. Thus, studies evaluating associations between measured ambient PAH levels and health outcomes are infrequent (Farmer et al. 2003, Taioli et al. 2007). No epidemiological investigations have used environmental PAH monitoring data in conjunction with a geographic information system. Some epidemiologic studies have used personal monitoring of PAH exposures (e.g. Perera et al. 2009). PM and some other traffic-related pollutants are commonly monitored and may be considered surrogates of ambient PAH or

For traffic exposure in particular, previous epidemiological investigations have generally employed relatively crude measures such as distance to major roads or traffic density (Nie et al. 2007, Savitz and Feingold 1989). These measures are often based on self-reported data on "street type, traffic intensity, frequencies of traffic jams at the home address, or proximity of the home to major roads," "air pollution annoyance scores," or "traffic annoyance scores" (Heinrich et al. 2005) (Boothe and Shendell 2008). Other investigations have evaluated less subjective traffic exposure surrogates, such as "traffic counts on major roads, air pollution data from municipal monitoring sites, files on traffic counts at home address, geographic information system (GIS) based assessment of distance from subject's home to a major road, spatial air pollution measurements, interpolated concentrations derived from monitoring network data...dispersion models...and finally complex regression models using measured air pollution data and GIS data on traffic count, distance to major roads, and population density" (Heinrich et al. 2005) (Boothe and Shendell 2008).

One previous study has taken excess traffic exhaust emissions at intersections and during engine warm-up into account (Nie et al. 2007). Otherwise, none of these methods account for elevated emissions at intersections or shortly following engine start-up (Beyea et al. 2006, Carr et al. 2002, Nie et al. 2007). Many studies do not validate their exposure estimates. Research indicates that modeled traffic exposure estimates predict pollutant concentrations more effectively than cruder traffic exposure proxies such as "surrogate variables or...ambient monitoring data" (Brauer et al. 2003) (Heinrich et al. 2005).

Several studies have evaluated associations between traffic or air pollution exposure and breast cancer (Bonner et al. 2005, Crouse et al. 2010, Lewis-Michl et al. 1996, Nie et al. 2007,
Raaschou-Nielsen et al. 2011). One of these investigations calculated residential traffic and chemical facility density using a geographic information system divided into 5 km² grids and 1 km² grids, respectively (Lewis-Michl et al. 1996). Four other population-based case-control studies examined relations between more individualized air pollution exposure estimates and breast cancer risk (Bonner et al. 2005, Crouse et al. 2010, Nie et al. 2007, Raaschou-Nielsen et al. 2011). One investigation utilized a residential total suspended particulates (TSP) exposure estimate calculated from environmental monitor readings and subsequent "inverse distance squared weighed interpolation" (Bonner et al. 2005). Another study used a version of the historical residential traffic emissions exposure model originally developed for the Long Island Breast Cancer Study Project (Beyea et al. 2006, Nie et al. 2007). A recent hospital-based case-control investigation estimated ambient residential nitrogen dioxide concentrations as a marker for traffic pollution using a combination of pollutant monitoring data and land-use regression models (Crouse et al. 2009a,b, 2010). A cohort study evaluated associations between modeled long-term exposure to traffic-related nitrogen oxides and breast cancer risk (Raaschou-Nielsen et al. 2011). Nitrogen oxides are gaseous pollutants, and therefore do not reflect depletion and long-term transport phenomena relevant to carcinogenic particulate traffic air pollution (Beyea et al. 2006, IARC 2010). Finally, an ecological analysis reports associations between traffic-related pollutants and breast cancer incidence in the US (Wei et al. 2012).

**Vehicular Traffic and Overall Health Effects**

Numerous investigations report associations between a range of adverse health outcomes and (1) traffic-related pollutants such as benzene, particulate matter and carbon monoxide, (2) motor vehicle exhaust, and (3) traffic exposure surrogates such as residential proximity to major roads (Boothe and Shendell 2008). These health-related outcomes include adult and perinatal

Several studies have evaluated associations between traffic-related variables and female breast cancer, and all found some evidence of a positive association (Bonner et al. 2005, Crouse et al. 2010, Lewis-Michl et al. 1996, Nie et al. 2007, Raaschou-Nielsen et al. 2011). Occupational exposure to vehicular emissions has been linked to male breast cancer (Hansen 2000).

IARC classifies diesel exhaust as probably carcinogenic to humans and gasoline vehicle exhaust as possibly carcinogenic to humans (IARC 1989). In addition to PAHs (IARC 2010), confirmed human carcinogens in vehicle emissions include benzene (IARC 1987), formaldehyde (IARC 2006), and 1,3-butadiene (IARC 2008). All three carcinogens cause mammary tumors in laboratory animals following ingestion or inhalation (Hecht 2002, Huff et al. 1989, IARC 2006, 2008, Maltoni et al. 1988, 1989, Rudel et al. 2007). Benzene and formaldehyde have also been linked with breast cancer in occupational research (Cantor et al. 1995, IARC 2008, Petralia et al. 1999). In addition, the US EPA classifies the traffic-related pollutant acetaldehyde as a probable human carcinogen (US EPA 1994b) and IARC classifies this pollutant as a possible human carcinogen (IARC 1985a). A recent review noted that "because of their potency and high concentrations, most of the cancer risk [from traffic] has been attributed to benzene, 1,3-butadiene, and particle-bound PAHs” (Boothe and Shendell 2008).
Vehicular exhaust contains many highly correlated chemicals (Bostrom et al. 2002). Thus, it is not possible to definitively assess the contributions of individual traffic-related pollutants to health effects based on observational studies (Bostrom et al. 2002). Furthermore, the specific components of the traffic mixture can influence the health effects of any given chemical in that mixture. For example, "metabolism and bioactivation of individual PAH compounds are influenced by other PAHs" and transition metals found in particulate matter "can enhance the mutagenicity and carcinogenicity of other carcinogens such as PAHs" (Mehta et al. 2008). However, despite being correlated, traffic-related pollutants do not have identical distributions. For example, in the Long Island Breast Cancer Study Project (LIBCSP), residential traffic PAH exposure estimates were modeled using PAH-specific vehicle emissions data and PAH-specific dispersion parameters, and were validated and calibrated against residential soil PAH levels (Beyea et al. 2006). As an illustrative example, the benzo[a]pyrene geographic model used in the LIBCSP needed to be altered in order to predict carbon monoxide (CO) levels for a validation exercise. Specifically, the carbon monoxide model excluded "all depletion phenomena, because deposition, washout, and photo decay are negligible in the case of CO" (Beyea et al. 2006). Due to their "potency and high concentrations" in vehicular exhaust, particulate-phase PAHs are considered among the most carcinogenic traffic-related chemicals (Boothe and Shendell 2008). Benzo[a]pyrene is one of the most carcinogenic PAHs and is representative of PAH exposure from vehicle exhaust (Beyea et al. 2008, Bostrom et al. 2002, Fertmann et al. 2002).

**Associations of Traffic and Air Pollution with Breast Cancer**

Studies examining relations between air pollution and breast cancer are sparse. A case-control study conducted on Long Island (Nassau and Suffolk counties) reported a positive, but
not statistically significant, association between high residential traffic density in 5 km² grids and breast cancer (Lewis-Michl et al. 1996). This association was limited to participants residing in Nassau County. The study also reported elevated and statistically significant breast cancer odds among women living near chemical facilities, using a geographic information system divided into 1 km² grids (Lewis-Michl et al. 1996). Participants had resided on Long Island for at least 20 years.

Another population-based case-control study reported a positive relation between residential exposure to TSP at birth and postmenopausal breast cancer (OR = 2.42; 95% CI: 0.97, 6.09 for the highest vs. the lowest quartile, p for trend = 0.01; Bonner et al. 2005). This result was supported when evaluating TSP exposure at birth as a continuous variable (OR = 1.20; 95% CI: 1.04, 1.38 for every 30 μg/m³ increase in TSP), and by spline regression analysis (Bonner et al. 2005). Cumulative TSP exposure was estimated using data on residential TSP concentrations at birth, menarche, first birth, and 10 and 20 years prior to the case-control interview. This measure was also positively associated with postmenopausal breast cancer (OR = 3.5; 95% CI: 1.4, 8.9 for the highest quartile, p for trend = 0.03; Bonner et al. 2005). TSP was considered a surrogate for ambient PAH exposure in this investigation. TSP levels correlated well with ambient benzo[a]pyrene measurements during the one year in which both pollutants were monitored simultaneously in the study region (r = 0.90; Bonner et al. 2005).

The third relevant investigation, another population-based case-control study, reported positive associations between modeled residential exposure to traffic emissions (specifically, the model was adapted from the LIBCSP and estimated exposure to traffic benzo[a]pyrene) at the age of menarche and premenopausal breast cancer (OR = 2.07; 95% CI: 0.91 – 4.72 for the highest quartile; p for trend = 0.03), and between residential exposure to traffic emissions at the
age of first birth and postmenopausal breast cancer (OR = 2.58; 95% CI: 1.15, 5.83 for the highest quartile, p for trend = 0.19) (Nie et al. 2007). Upon stratifying by smoking history (ever/never), both associations were observed only among non-smokers. P-values for interaction were 0.01 and 0.06, respectively (Nie et al. 2007). Results did not differ with respect to tumor hormone receptor (ER/PR) status (Nie et al. 2007). Further research suggested that these associations may be limited to women with the \textit{GSTM1} null genotype (OR for the association between traffic exposure at menarche and premenopausal breast cancer = 4.64, 95% CI: 0.98, 21.94 for the highest vs. the lowest quartile, p for trend = 0.01; OR for the association between exposure at first birth and postmenopausal breast cancer = 3.27, 95% CI: 0.99-10.84, p for trend = 0.02) (Nie et al. 2005). Risk was reported as not elevated among women with the \textit{GSTM1} wild-type genotype, and variation in \textit{GSTM1} was not independently associated with breast cancer (though relevant effect estimates and confidence intervals were not reported) (Nie et al. 2005). No other variants were evaluated. \textit{GSTM1} is an important detoxification and antioxidative defense enzyme (Reszka et al. 2006).

A recent hospital-based case-control study found associations between the traffic-related pollutant nitrogen dioxide and postmenopausal breast cancer (OR = 1.31; 95% CI: 1.00-1.71 for each 5 ppb pollutant increase in the year 1996) (Crouse et al. 2010). The study evaluated 7 different regression models (for the year 2006, and 2 models each utilizing estimates from somewhat different exposure assessment methods for the years 1985, 1996, and for the mean of 1985 and 1996) (Crouse et al. 2010). No other effect estimates reached statistical significance. However, all effect estimates were fairly consistent in that they ranged between OR = 1.17 and OR = 1.36 (Crouse et al. 2010).

A large cohort study examined the association between modeled residential exposure to
nitrogen dioxide concentrations as a marker of traffic exposure and breast cancer incidence
(Raaschou-Nielsen et al. 2011). Age-adjusted models showed a positive association between
nitrogen oxides and breast cancer (incidence rate ratio [IRR] = 1.39, 95% CI: 1.09, 1.77 per 100
μg/m³ NOx), which was attenuated to some extent after adjustment for other covariates (IRR =
1.16, 95% CI: 0.89, 1.51).

Finally, a recent ecological analysis reported associations between traffic-related

PAH AND BREAST CANCER: PROPOSED MECHANISMS

**PAHs as Animal Mammary Carcinogens**

Several PAHs, including benzo[a]pyrene, reliably induce mammary tumors in laboratory
rodents (el-Bayoumy 1992, IARC 2010). This is demonstrated by experimental studies in which
PAHs (for example, benzo[a]pyrene, dibenzo[a,h]pyrene, dibenzo[a,i]pyrene,
dibenzo[a,l]pyrene) were administered orally (el-Bayoumy et al. 1995, Huggins and Yang 1962),
by direct injection to the mammary gland (Cavalieri et al. 1989, 1991, Hecht et al. 2002),
through skin painting (Ranadive and Karande 1963), and through intracolonic instillation
(Anderson 1983, IARC 2010). Two PAH compounds are in fact used as model mammary
carcinogens for experimental studies in Sprague-Dawley rats (Santodonato 1997). No
experimental studies examining associations between inhaled PAHs or PAH sources and
malignant mammary tumors were identified in the literature. One investigation evaluated and
reported positive associations between inhaled diesel exhaust and mammary adenomas and
fibromas (i.e. benign growths) among rats (Iwai et al. 1986).

**PAHs as Human Carcinogens**

Benzo[a]pyrene, several PAH-related occupational exposures (aluminum production,
roofing, paving, chimney sweeping, coal-tar distillation, coke production, coal gasification), and tobacco smoke are classified as known human carcinogens by IARC (IARC 1984, 1985b, 2004, 2010, Straif et al. 2005). Benzo[a]pyrene was classified as a human carcinogen “based on sufficient evidence in animals and strong evidence that the mechanisms of carcinogenesis in animals also operate in exposed human beings” (Straif et al. 2005). PAH-containing occupational mixtures were classified as carcinogenic to humans based on results of both occupational studies and animal research (Straif et al. 2005).

Other individual PAHs (for example, benz[a]anthracene, indeno[1,2,3-cd]pyrene, benzo[b]fluoranthene, dibenz[a,h]anthracene, benzo[k]fluoranthene, chrysene) and PAH-containing mixtures, including diesel engine exhaust and gasoline engine exhaust, are classified as probable or possible human carcinogens by health and environmental agencies based on animal research, epidemiologic/occupational studies, or mechanistic data (Bostrom et al. 2002, CA EPA 1999, IARC 1989, 2010, Minesotta Department of Health, NTP 2011, US EPA 2011a). It should be noted that (1) some PAH-containing mixtures, such as vehicular exhaust and cigarette smoke, contain confirmed, probable or possible human carcinogens other than PAHs (Bostrom et al. 2002), and (2) the above classifications are primarily based on tumorigenesis in tissues and organs other than the breast (i.e. primarily lung and skin cancers) (IARC 2010). There has been a paucity of research on associations between PAHs or PAH exposure sources and breast cancer risk.

**Initiation, Promotion, Progression, and Dose-Response**

Carcinogenesis for epithelial cancers such as breast cancer is conceptually divided into three broad stages: initiation, promotion, and progression (Bostrom et al. 2002, Pitot and Dragan 1996). Initiation involves carcinogen or metabolism-mediated mutations in tumor suppressor
genes or proto-oncogenes (Bostrom et al. 2002, Kinzler and Vogelstein 1998). Promotion involves proliferation of initiated cells (Bostrom et al. 2002). Progression is an irreversible process by which a cell gains further mutations and genetic instability and completes the process of attaining "malignancy and autonomous cell growth" (Bostrom et al. 2002). In order for cancer to occur, a tumor must also evade host defenses, such as immunologic and DNA repair processes (Morris and Seifter 1992, Nasca and Pastides 2008).

Based on in vivo and in vitro experimental investigations, several PAHs including benzo[a]pyrene are known to be complete carcinogens that can act as initiators, promoters, and progressors through a combination of genotoxic and estrogenic mechanisms (Bostrom et al. 2002, Cerutti et al. 1977, Morris and Seifter 1992). PAHs may also interact with other inherited or lifestyle/environmental factors in exerting a carcinogenic effect (Bostrom et al. 2002). Individual PAHs differ in metabolic and mutagenic/carcinogenic properties (IARC 2010). More structurally complex PAHs tend to have higher mutagenic and carcinogenic capacity (Bostrom et al. 2002).

Compounds acting via initiator or promoter mechanisms tend to have different dose-response relationships with tumor formation. Pure promoters generally show an approximately S-shaped dose-response curve, whereas purely genotoxic chemicals tend to show a more linear dose-response relationship (Bostrom et al. 2002). Most animal studies yield non-linear dose-response relationships between PAHs and carcinogenesis "with an upward rise at high doses" (Bostrom et al. 2002), which can be linearized upon simultaneous administration of another promoter (Bostrom et al. 2002, Burns et al. 1983). Animal research usually employs very high PAH doses, however, whereas humans are generally environmentally exposed to relatively low PAH levels (Bostrom et al. 2002). A review noted
that "at the present state of knowledge, risk estimation for PAHs at low exposure levels should be based on the assumption of linear dose-response relationships despite the nonlinear responses often seen for high doses in animal tests" (Bostrom et al. 2002).

**PAH Properties Relevant to Breast Carcinogenesis**

PAHs are lipophilic, meaning that they are stored and can accumulate in adipose tissue, including in the breast (Morris and Seifter 1992, Obana et al. 1981, Terry and Rohan 2002). PAHs exhibit estrogenic and anti-estrogenic (Santodonato 1997), inflammatory (Jeng et al. 2010, Schober et al. 2007), pro-oxidant (Farmer et al. 2003, Jeng et al. 2010) and genotoxic (Farmer et al. 2003) properties that may be relevant to breast carcinogenesis (Gammon and Santella 2008).

Estrogens are endogenous steroid hormones that regulate breast growth and differentiation during adolescence and pregnancy and stimulate cell division in healthy or malignant estrogen receptor-positive breast tissue (Nasca and Pastides 2008). PAHs are similar to estrogens with respect to structure, metabolism and transport (Morris and Seifter 1992). Estrogenic effects of PAHs may be especially relevant when a woman's endogenous estrogen levels are low, such as after menopause (Brody and Rudel 2003). Exposure to environmental estrogens may contribute to breast carcinogenesis through (1) increased likelihood of spontaneous DNA replication errors due to higher levels of cell proliferation, (2) increased likelihood of DNA damage going uncorrected by DNA repair mechanisms due to higher levels of cell proliferation, and (3) promoting the growth of existing small tumors or initiated cells (Brody and Rudel 2003, Nasca and Pastides 2008). Benzo[a]pyrene increases proliferation of both untransformed human mammary epithelial cells and human breast carcinoma cells in *in vitro* experimental research (IARC 2010, Tannheimer et al. 1997, 1998, Tsai et al. 2004b).

PAHs may also influence cancer risk through immunotoxic or inflammatory mechanisms
(Bostrom et al. 2002, IARC 2010, Jeng et al. 2010, Schober et al. 2007, Straif et al. 2005). Since a well-functioning immune system can help protect against tumor formation, reduced immune function could adversely affect cancer risk (IARC 2010). High-level PAH exposure has been linked to immunotoxic endpoints in animal studies within many species, in human cell lines, and in mechanistic studies (Davila et al. 1995, IARC 2010, Ramesh et al. 2004). There is also epidemiological and clinical evidence that environmental exposure to PAHs may affect immune function in humans, but research is as yet too sparse to draw a definitive conclusion (IARC 2010). Similarly, chronic inflammation is linked to carcinogenesis (Fitzpatrick 2001), and PAHs are associated with induction of inflammatory activity in experimental research (Jeng et al. 2010, Schober et al. 2007). It should be noted that because PAHs are often bound to particles, “there is considerable potential for covariance with an inflammatory response that is induced by the carrier particles alone” (IARC 2010).

Some PAHs have a high affinity to the cytosolic aryl hydrocarbon (Ah) receptor, which is distributed throughout the body, including in breast tissue (IARC 2010, Safe 2001). Activation of this receptor leads to upregulation of genes relevant to cancer promotion, such as those involved in growth and differentiation (Bostrom et al. 2002). In animal studies, Ah receptor activation by aromatic hydrocarbons induces endocrine disruptor effects, cell proliferation and tumor promotion (Bostrom et al. 2002). The Ah receptor is also involved in the induction of enzymes which metabolize PAHs to electrophilic genotoxic intermediates (specifically, cytochrome P450s) (Bostrom et al. 2002, IARC 2010, Santodonato 1997). In summary, “several of the biological effects of PAHs, such as enzyme induction of xenobiotic metabolizing enzymes, immunosuppression, teratogenicity and carcinogenicity, are thought to be mediated by activation of the aryl hydrocarbon receptor” (IARC 2010).
Finally, PAHs may contribute to breast carcinogenesis through induction of DNA damage, both by direct interaction with DNA and via oxidative stress-related mechanisms (Harvey et al. 1996). This is discussed in much more detail below.

**PAH Exposure, Absorption and Metabolism**

PAHs are rapidly absorbed into the human body following inhalation, ingestion, or skin contact (IARC 2010), and are released into the general circulation following all three exposure routes (IARC 2010, Modica et al. 1983, Rees et al. 1971, Withey et al. 1993a,b). Because this document focuses primarily on the carcinogenic effects of ambient PAH pollution, the following review mainly concentrates on absorption and metabolism of PAHs following inhalation exposure.

It should be noted that gastrointestinal absorption may also occur following inhalation exposure, due to mucociliary clearance and release of PAH metabolites following hepatobiliary processing (IARC 2010, Sun et al. 1982). Ingested PAHs are absorbed by diffusion through the gastrointestinal tract epithelium (IARC 2010, O'Neill et al. 1991). Gastrointestinal absorption and metabolism of PAHs is influenced by PAH size, other dietary factors (i.e. drugs, foods or nutrients, and environmental contaminants), gastric hormones, and bile salts (IARC 2010, Ramesh et al. 2004).

Ambient PAHs are absorbed through the airway epithelium upon inhalation, and then distributed to the circulation and to other tissues, where PAHs are metabolized through activation and detoxification pathways (Bostrom et al. 2002, Castaño-Vinyals et al. 2004, Gammon and Santella 2008). Metabolism of PAHs can occur at the site of entry as well, though in practice the lung shows limited metabolic activity (Bostrom et al. 2002). Absorption of inhaled PAHs is influenced by (1) the degree of lipophilicity, such that more lipophilic compounds diffuse at
lower rates (Bostrom et al. 2002), (2) whether the PAH has adhered to a particle, with slower clearance of particle-associated PAHs (Brzeznicki et al. 1997, Castaño-Vinyals et al. 2004, Creasia et al. 1976, Sun et al. 1982), (3) the location of PAH deposition on the airway epithelium (PAHs diffuse more quickly from the alveolar epithelium than from the epithelia of conducting and bronchial airways) (Bostrom et al. 2002, Gerde and Scott 2001, IARC 2010), (4) the size of the PAH molecule, as PAHs with fewer rings are absorbed more quickly and efficiently (IARC 2010), and (5) if particle associated, the size of the particle to which a PAH is bound and the rate at which the PAH dissociates from the particle (IARC 2010).

Inhaled PAHs are released into the general circulation following absorption and diffusion through the airway epithelium, basement membrane, and finally the endothelium (Bostrom et al. 2002). This process can occur relatively quickly (IARC 2010). For example, an experimental study in rats was able to measure radiolabelled benzo[a]pyrene and its metabolites in the blood within 95 minutes of inhalation exposure (IARC 2010, Withey et al. 1993b). Upon entering the general circulation, PAHs are “widely distributed to most organs and tissues...with some preferential distribution to or retention in fatty tissues” (IARC 2010). The breast contains a high proportion of adipose tissue (Yaffe et al. 2009). PAHs are lipophilic and "can accumulate in adipose tissue to be released over time and redistributed to the breast epithelial cells” (Rundle et al. 2000a) (Obana et al. 1981).

The majority of PAH metabolism occurs in the liver, regardless of the route of exposure and entry (Bostrom et al. 2002, IARC 2010, Ramesh et al. 2004). Resulting metabolites (including DNA-reactive metabolites) can be excreted via the biliary or renal systems or can be released into the general circulation and distributed to extra-hepatic tissues (Bostrom et al. 2002, Ramesh et al. 2004). PAHs are metabolized and activated to reactive carcinogenic intermediates.

PAHs are generally carcinogenic and chemically reactive only upon metabolic activation to charged electrophilic intermediates via Phase I pathways (Cavalieri and Rogan 1985, Gammon and Santella 2008, Hecht 1999, IARC 2010, Morris and Seifter 1992). Many PAH metabolites, including DNA-reactive metabolites (IARC 2010), are produced during the course of Phase I metabolism (IARC 2010, Shimada 2006). However, PAHs are ultimately mainly activated to the following DNA-reactive/carcinogenic metabolites: diol epoxides, radical cations, and PAH orthoquinones (Bostrom et al. 2002, IARC 2010, Jerina et al. 1991, Ramesh et al. 2004, Straif et al. 2005). Metabolism of PAHs in relation to breast cancer risk is illustrated in Figure A1.1. Experimental \textit{in vitro} studies show that human breast epithelial tissue has the ability to metabolize PAH to their ultimate mutagenic/carcinogenic moieties (Calaf and Russo 1993, Eldridge et al. 1992, Mane et al. 1990)" (Rundle et al. 2000a) (MacNicoll et al. 1980, Terry and Rohan 2002).


Reactive PAH metabolites from Phase I metabolism are detoxified to more polar and hydrophilic compounds through Phase II metabolic pathways (Shimada 2006). For example, glutathione S-transferases such as GSTM1, GSTP1 and GSTT1 detoxify activated PAH metabolites through conjugation with glutathione (Bostrom et al. 2002, Ramesh et al. 2004, Straif et al. 2005, Terry and Rohan 2002). Other detoxifying enzymes include UDP-glucuronosyltransferases and sulfotransferases (Bostrom et al. 2002, IARC 2010, Straif et al. 2005, Terry and Rohan 2002). PAHs are thus commonly detoxified to metabolites including sulfate conjugates, glutathione conjugates, and glucuronides (Bostrom et al. 2002, Castaño-Vinyals et al. 2004). The products of Phase II metabolism are excreted, sometimes following further metabolic processing (Bostrom et al. 2002); this process "is relatively efficient
because of the wide distribution of enzymes that transform PAHs into polar metabolites" (ATSDR 2009).

PAHs can induce PAH-activating cytochrome P450 enzymes via the aryl hydrocarbon receptor (IARC 2010, Ramesh et al. 2004; Shimada 2006). Similarly, PAHs and PAH metabolites can induce prostaglandin H synthase (Kelley et al. 1997). In contrast, PAHs can also inhibit cytochrome P450 activity (Shimada 2006) and induce glutathione S-transferases (Pushparajah et al. 2008). Thus, PAH exposure levels and conditions can “influence the balance of Phase I and Phase II enzymes, which can determine whether or not a toxic cellular response occurs” (IARC 2010).

Genetic variation in Phase I and Phase II pathway genes, which could influence the expression of the corresponding enzymes, is hypothesized to partially underlie differential cancer susceptibility upon PAH exposure (IARC 2010). However, relevant epidemiological studies have been inconsistent (e.g. McCarty et al. 2009, Rundle et al. 2000b). This may be due to insufficient power in some investigations or differences in participant characteristics such as genotype combinations, race, "age, sex, pathological and physiological conditions, and exposure to environmental pollutants" (Shimada 2006) across study populations (IARC 2010, Shimada 2006).

**PAHs and DNA Adducts**

Given sufficiently high levels of PAH exposure or inadequate host detoxification, DNA-reactive PAH metabolites bind to DNA and form adducts (Dipple et al. 1984, Gammon and Santella 2008, Harvey et al. 1991, 1996). PAH metabolites preferentially bind to guanine residues (Braithwaite et al. 1998), and will almost always bind to either guanine or adenine (IARC 2010). PAH-DNA adducts are found in human breast tissue and human breast milk

PAH-DNA adducts reflect both PAH exposure levels and host response (Gammon and Santella 2008, Santella 1999). As demonstrated by *in vitro* and *in vivo* experimentation, bulky DNA adducts are generally repaired through nucleotide excision repair (NER) pathways (Braithwaite et al. 1998, Chakravarti et al. 2008, Wei et al. 1995), though base excision repair (BER) pathways may also contribute to bulky adduct repair (Braithwaite et al. 1998). Apurinic sites left by depurinating DNA adducts are generally repaired via BER pathways (Braithwaite et al. 1998, Chakravarti et al. 2008). DNA repair processes may be inadequate to fully correct the damage (Braithwaite et al. 1998), in which case PAH-induced adducts could ultimately lead to somatic mutations in cancer-related genes (Gammon and Santella 2008, Harvey et al. 1996). NER is a slower pathway than BER, so repair may be less efficient for stable than for depurinating adducts (Braithwaite et al. 1998, Leavitt et al. 2008).

Diol epoxide-derived adducts tend to be stable, whereas radical cations form unstable adducts “which spontaneously depurinate, leaving potentially promutagenic apurinic (AP) sites” (Braithwaite et al. 1998, Cavalieri and Rogan 1995, 1998, Devanesan et al. 1996, Drinkwater et al. 1980, IARC 2010, Leavitt et al. 2008, Ochi et al. 1986). Ortho-quinones commonly lead to both depurinating and stable adducts (Burczynski and Penning 2000, McCoull et al. 1999). AP sites from depurinating adducts are routinely formed as a result of normal cellular metabolism and are repaired quickly and efficiently (Leavitt et al. 2008). DNA damage is more likely to go unrepaired at higher levels of AP site generation (Leavitt et al. 2008). It is unclear whether stable
or depurinating adducts contribute more strongly to carcinogenesis (Shimada 2006).

In addition to individual differences in DNA repair capacity and adduct stability, adduct removal may be influenced by the surrounding bases (Bostrom et al. 2002, Hess et al. 1997) and by whether the adduct is found on a coding or non-coding DNA strand (Bostrom et al. 2002, Jernstrom and Graslund 1994, Mellon et al. 1987). Adduct formation and persistence is also influenced by individual differences in cell cycle control, PAH exposure levels, PAH activation, detoxification and absorption, and lifestyle or environmental factors. For example, enzymes relevant to PAH metabolism may be induced by medicines or environmental exposures, whereas other substances may act in a chemopreventive capacity (Eder 1999, Gammon and Santella 2008, IARC 2010, Santella 1999).

PAHs and Reactive Oxygen Species

PAHs may also increase cancer risk through induction of reactive oxygen species (ROS) (Bolton et al. 2000, Flowers et al. 1997, IARC 2010, Singh et al. 2007, Yu et al. 2002). PAHs influence ROS formation via (1) redox cycling of PAH quinone metabolites (Harvey et al. 1996, IARC 2010, Penning et al. 1999, Singh et al. 2007), (2) peroxidase reactions (e.g. involving myeloperoxidase; Gammon and Santella 2008, IARC 2010, Podrez et al. 2000, Singh et al. 2007), and (3) inducing inflammation (Ferguson 2010, Jeng et al. 2010, Schober et al. 2007). Even “small amounts of quinone metabolites may result in generation of high ratios of reactive oxygen species,” such as hydroxyl radicals and superoxide anion (Flowers-Geary et al. 1993, Harvey et al. 1996, Penning et al. 1996).

Antioxidants can counteract PAH-induced ROS. Superoxide dismutase and GST enzymes are examples of endogenous antioxidants (Gammon and Santella 2008, Scandalios 2005). Exogenous antioxidants include beta-carotene and vitamins C and E (Klaunig and
Oxidative stress occurs when ROS are not balanced by antioxidants (Klaunig and Kamendulis 2004).

ROS damage DNA both directly and by forming DNA-damaging lipid peroxidation products (Farmer et al. 2003, Gammon and Santella 2008, IARC 2010). Over 20 types of oxidative DNA damage have been identified (Cooke et al. 2003, Ferguson 2010). The most common oxidative stress-induced DNA lesion is 8-oxodeoxyguanosine (Boiteux and Radicella 1999, Piette 1991, Rossner et al. 2006), which causes somatic mutations, especially GC- TA transversions (Boiteux and Radicella 1999, Grollman and Moriya 1993).

Oxidative DNA damage is caused by both environmental exposures and normal cellular processes (Boiteux and Radicella 1999). At background DNA damage levels, oxidized bases are efficiently and constitutively repaired (Leavitt et al. 2008). However, uncorrected oxidative damage commonly occurs, especially at high environmental exposure and oxidative stress levels (Boiteux and Radicella 1999, Ferguson 2010). When DNA repair pathways are inadequate to correct oxidative damage from PAHs, the resulting genetic damage persists (Boiteux and Radicella 1999, Grollman and Moriya 1993, Gammon and Santella 2008). The extent of PAH-induced oxidative damage may differ between individuals based on variations in PAH exposure level and frequency, activation and detoxification processes, DNA repair capacity and other defensive mechanisms, or antioxidant intake (IARC 2010, Singh et al. 2007). Oxidative DNA damage is repaired mainly through the BER pathway (Goode et al. 2002), though NER can play a role as well (Gros et al. 2002).

Detoxification of PAH-induced ROS employs limited antioxidant resources, which may have otherwise protected against oxidative damage from other sources (Morris and Seifter 1992). ROS can also lead to “electrophilic and pro-oxidant signals that may affect cell growth” (IARC
Finally, recent research suggests that oxidative stress may lead to inhibition of NER activity, including repair of PAH-DNA adducts (Ferguson 2010, Mehta et al. 2008). Oxidative stress is strongly linked to carcinogenesis in experimental research (Klaunig and Kamendulis 2004). Relevant studies are suggestive of an association between oxidative stress and breast cancer in humans (Li et al. 2001, Ray et al. 2000, Wang et al. 1996). However, there is as yet insufficient evidence to conclusively confirm this relationship (Shen et al. 2009).

**DNA Damage, PAH-Induced Carcinogenesis and the Role of DNA Repair**

To summarize, several major mechanisms are proposed to underlie PAH-induced carcinogenesis (Harvey et al. 1996, IARC 2010), including metabolic formation of diol epoxides, quinones and radical cations, as well as induction of oxidative stress. Metabolic formation of diol epoxides is the most widely accepted and likely most important proposed mechanism (Harvey et al. 1996). Each of these processes may damage DNA and eventually lead to critical mutations in tumor suppressor genes or proto-oncogenes (Harvey et al. 1996, IARC 2010). These proposed “mechanisms are not mutually exclusive, and all may play a role in the induction of cancer” (Harvey et al. 1996) (IARC 2010).

In addition to detoxification of reactive metabolites, which may prevent DNA damage from occurring in the first place, the effects of existing DNA damage may be mitigated by a number of defensive mechanisms, including apoptosis, cell cycle arrest, and DNA repair (Goode et al. 2002, Harvey et al. 1996). If these mechanisms fail or are insufficient, resulting uncorrected DNA damage can contribute to cancer risk through eventual formation of somatic mutations in tumor suppressor genes (such as p53) or proto-oncogenes (such as ras) during replication (Bigger et al. 1990, 1991, Braithwaite et al. 1998, Chakravarti et al. 2008, Glatt et al. 1989, Harvey et al. 1996, Straif et al. 2005). PAHs induce mutations in proto-oncogenes and

Because persistent mutations in critical genes can lead to genetic instability and carcinogenesis (Berwick and Vineis 2000, Neumann et al. 2005), polymorphic genes in DNA repair pathways are standard candidate cancer susceptibility genes (Gammon and Santella 2008). Certain DNA repair gene polymorphisms (for example, in \textit{OGG1} and \textit{XRCC1}) are consistently associated with overall cancer risk in epidemiologic studies (Goode et al. 2002). Reduced DNA repair (e.g. as measured by host-cell reactivation or benzo[a]pyrene diol-epoxide [BPDE] sensitivity assays) is consistently associated with risk of breast (Helzlsouer et al. 1996, Kennedy et al. 2005, Kosti et al. 2010, Motykiewicz et al. 2002, Parshad et al. 1996, Ramos et al. 2004, Shi et al. 2004, Xiong et al. 2001) and other cancers (Berwick and Vineis 2000, Lockett et al. 2005). Some confirmed breast cancer susceptibility genes are involved in DNA repair pathways (e.g. \textit{BRCA1}, \textit{BRCA2}; Kosti et al. 2010, Weber and Nathanson 2000).

Polymorphisms in DNA repair genes may modify effects of environmental exposures on carcinogenesis (Berwick and Vineis 2000, Gammon and Santella 2008), even without otherwise exhibiting a main effect on breast cancer risk. It is hypothesized that associations between some DNA repair variants and cancer risk "may be apparent only in the presence of DNA-damaging agents such as tobacco smoke" (Goode et al. 2002). Several investigations have been suggestive
of interactions between PAH-DNA adducts or cigarette smoking (a PAH source) and DNA repair polymorphisms or capacity with respect to breast cancer risk (Crew et al. 2007, Kosti et al. 2010, Mechanic et al. 2006, Metsola et al. 2005, Pachkowski et al. 2006, Shen et al. 2005a, Shore et al. 2008, Terry et al. 2004).

DNA REPAIR MECHANISMS: BIOLOGY AND EPIDEMIOLOGY


MMR corrects DNA polymerase-mediated replication errors (Goode et al. 2002). This process operates through enzymes including MSH6, MLH1, MSH2, and PMS2 (Goode et al. 2002). Double-strand break repair, the only type of DNA repair in which no undamaged template is available, corrects DNA damage due to either replication errors or exogenous exposures such as ionizing radiation (Braithwaite et al. 1999, Goode et al. 2002). Double-strand break repair can proceed through the homologous recombination (HR) pathway or the nonhomologous end-joining repair (NHEJ) pathway (Goode et al. 2002), and involves enzymes including BRCA1, BRCA2, and XRCC3 (Goode et al. 2002). The HR pathway "is a process by which double-strand DNA breaks are repaired through the alignment of homologous sequences of DNA," whereas "the NHEJ repair pathway involves direct ligation of the two double strand break ends" (Kiyohara and Yoshimasu 2007).

NER and BER pathways remove damaged nucleotide bases and are therefore important to correcting PAH-induced DNA damage (Braithwaite et al. 1999). These pathways are described in more detail below, in general and with respect to PAH damage and breast carcinogenesis. In vitro and in vivo laboratory investigations demonstrate that PAH-DNA adducts are repaired by
NER (Braithwaite et al. 1998, Gunz et al. 1996, Koostra 1982, van Houten 1990, Venkatachalam et al. 1995). BER repairs PAH-induced oxidative DNA damage (Braithwaite et al. 1998, Goode et al. 2002) and may also play a role in repairing PAH-DNA adducts (Blakey and Douglas 1990, Braithwaite et al. 1998, Day et al. 1978). Many genes encoding DNA repair enzymes are polymorphic. These “polymorphisms can exist in the promoter regions, introns and exons of a given gene and therefore can affect transcription, mRNA processing, functional activity or enzyme stability, respectively” (IARC 2010).

The impact of DNA repair gene polymorphisms on DNA repair function is still largely unknown (Mechanic et al. 2006). These functional effects are challenging to study because of (1) assay differences between studies, (2) variations in study population characteristics (for example, DNA repair capacity may be affected by factors such as age and smoking status), and (3) difficulty detecting "subtle difference in [DNA repair capacity] by studying only one or two SNPs in a very complex pathway" (Shen et al. 2006). It is likely that “end-points of exposure to PAHs such as metabolic profile, DNA adducts...and tumor incidence may result from multigene interactions” (IARC 2010).

Similarly, studies of associations between DNA repair genes and cancer risk are often hampered by insufficient sample size, thereby increasing the likelihood of both false-negative and false-positive results (Rothman and Greenland 1998), and by other methodological issues including inappropriate genotypic groupings, confounder control strategies, or control group selection (Goode et al. 2002). Few investigations have assessed haplotypes or gene-gene interactions with respect to breast or other cancer risk. Such investigations may yield important information given sufficient statistical power (Gammon and Santella 2008, Goode et al. 2002, Khoury et al. 1993, Kiyohara and Yoshimasu 2007).
The following sections include a review of the literature regarding associations between
selected BER and NER genes (\textit{XPA}, \textit{XPD}, \textit{XPF}, \textit{XPG}, \textit{ERCC1}, \textit{XRCC1}, and \textit{OGG1}) and DNA
repair capacity or DNA damage levels, with a special focus on PAH-related DNA damage.

Benzo[a]pyrene exposure is associated with micronuclei, sister chromatid exchanges, cell
transformation, chromosomal aberrations, somatic mutations and DNA strand breaks in
experimental research (IARC 1983, IARC 2010). Unless otherwise noted, the laboratory
experiments described below examined \textit{in vitro} human cells, rather than animal models. The
following text also gives an overview of the NER and BER pathways, and summarizes
associations between selected DNA repair variants and breast cancer.

\textit{Nucleotide Excision Repair (NER)}

\textbf{Overview of the NER Pathway}

The NER pathway repairs DNA damage from bulky lesions that alter the shape of the
DNA helix (Greenblatt et al. 1994), including thymine dimers, ultraviolet-induced lesions, cross-
links, some ROS-induced damage, and bulky adducts, including PAH-DNA adducts (Braithwaite
al. 2006, Reardon et al. 1997). This repair process is unique in repairing such a wide variety of
DNA damage (Braithwaite et al. 1998, Friedberg et al. 1995). The NER pathway is divided into
two subpathways: transcription-coupled repair (TCR) and global genome repair (GGR)
(Hanawalt 2002, Kiyohara and Yoshimasu 2007). TCR "preferentially removes DNA damage
that blocks ongoing transcription in the transcribed DNA strand of active genes," whereas GGR
"removes lesions throughout the genome, including those from the nontranscribed strand in the
active gene" (Hanawalt 2002, Kiyohara and Yoshimasu 2007). These subpathways differ with
respect to the DNA damage recognition process (Wang et al. 2010).
NER is a complex multistep process involving at least 20 to 30 proteins (Braithwaite et al. 1998, Wood 1997), including from the xeroderma pigmentosum (e.g. XPA, XPC, XPD, XPF, XPG) and excision repair cross-complementing (e.g. ERCC1, ERCC6) groups (Sancar et al. 2004). The first step of the NER process is damage recognition and binding by a complex of proteins (Goode et al. 2002, Crew et al. 2007, Kiyohara and Yoshimasu 2007), which differs for the GGR (e.g. XPA, XPC) and TCR (e.g. ERCC6, ERCC8) subpathways (Lindahl and Wood 1999, Wang et al. 2010). Next, the TFIIH transcription factor complex of proteins, including the crucial adenosine triphosphate (ATP) dependent helicase XPD, unwinds DNA at the lesion site (de Boer and Hoijmakers 2000, Goode et al. 2002, Kiyohara and Yoshimasu 2007). Another complex of proteins including ERCC1, XPF and XPG makes dual incisions around the site of the DNA damage and excises the damaged single-stranded nucleotide fragment, which is usually approximately 27-30 base pairs in length (Goode et al. 2002, Mechanic et al. 2006). New DNA is synthesized by DNA polymerases, and DNA ligation completes the NER process (Braithwaite et al. 1998, Friedberg et al. 1995, Goode et al. 2002).

Selected NER Genes and DNA Repair Phenotype

Selected NER polymorphisms of interest are those genotyped in the LIBCSP: *XPA* (-4A/G), *ERCC1* (8092C/A), *XPD* (Lys751Gln, Asp312Asn), *XPF* (Arg415Gln) and *XPG* (Asp1104His). Some of these polymorphisms alter the resulting protein's amino acid sequence (*XPD* Lys751Gln and Asp312Asn, *XPF* Arg415Gln, and *XPG* His1104Asp). Others are located outside of coding regions (*XPA* -4A/G and *ERCC1* 8092C/A). A recent report stated that "based upon the degree of conservation across species and the change in polarity, charge and protein structure caused by the substituted amino acid, the *XPD* codon 312, *XPD* codon 751, [and] *XPG* codon 1104...polymorphisms are predicted to have significant functional impact" (Mechanic et al. 2006) (Shen et al. 1998a).

Experimental studies report associations between the *XPA* 23A/G variant polymorphism and (1) reduced DNA repair capacity in "cells transfected with benzo[a]pyrene dioleopoxide (BPDE)-treated plasmid" (Wu et al. 2003) as well as (2) an increased number of BPDE-induced chromatid breaks in human lymphocytes (Lin et al. 2007). Another experimental study reported null results for the association between the *XPA* 62T/C polymorphism and DNA repair capacity measured by removal of induced BPDE-DNA adducts, though a weak positive association was observed with a multivariable model including variants in *XPA*, *XPC*, and *XPG* (Shen et al. 2006). An epidemiologic investigation reported a null association between *XPA* (23A/G) and PAH-DNA adducts among lung cancer cases (Zienolddiny et al. 2006). The *XPA* 23A/G polymorphism was reported to modify PAH-induced DNA damage in coke oven workers (Pavanello et al. 2005, Wang et al. 2010).

A polymorphism in *ERCC1* (8092C/A) was associated with BPDE-induced DNA adducts in one experimental study (Zhao et al. 2008). Other laboratory investigations reported null
results for the 8092C/A polymorphism with respect to number of BPDE-induced chromatid breaks (Lin et al. 2007) and DNA repair capacity measured by removal of induced BPDE-DNA adducts (Shen et al. 2006). Epidemiological studies have reported associations between variants in ERCC1 and chromosomal DNA damage in coke-oven workers (Cheng et al. 2007; 19007 T/C) as well as reduced PAH-DNA adduct levels among lung cancer cases (Zienolddiny et al. 2006; 8092C/A). A study conducted among coke-oven workers reported null associations between any of 4 polymorphisms in ERCC1 (including 19007 T/C, but not 8092C/A) and DNA damage measured by the Comet assay (Wang et al. 2010).

Epidemiological studies report null associations between variants in XPF and chromosomal damage measured by cytokinesis-block micronucleus frequency (Cheng et al. 2007; Ser835Ser) and PAH-DNA adducts among lung cancer cases (Zienolddiny et al. 2006; Pro379Ser, Arg415Gln). No in vitro studies examining variants in XPF and DNA repair capacity in human cells were identified.

One experimental study reported a positive association between the XPG Asp1104His variant polymorphism and DNA single strand breaks (Vodicka et al. 2004). Another laboratory investigation found that a multivariable model accounting for polymorphisms in XPA, XPC, and XPG was weakly predictive of BPDE-DNA adduct removal in lymphoblastoid cells, though XPG Asp1104His was not independently associated with BPDE-DNA adduct removal (Shen et al. 2006). A third experimental study reported null results for Asp1104His and BPDE-induced chromatid breaks (Lin et al. 2007). An epidemiological study found a negative association between the XPG His46His variant polymorphism and PAH-DNA adducts in the lung tissue of cancer patients (Zienolddiny et al. 2006). Other epidemiologic studies report null results with respect to chromosomal and DNA damage levels in coke-oven workers (Cheng et al. 2007 for
Asp1104His, Wang et al. 2010 for seven variants including Asp1140His), as well as for bladder tumor p53 mutation status (Ryk et al. 2006; Asp1104His).

The most widely studied NER polymorphisms with respect to DNA repair function are the Asp312Asn and Lys751Gln SNPs in XPD (Pabalan et al. 2010). Null associations between these variants and DNA repair capacity were reported in several controlled laboratory studies (Laine et al. 2007, Shen et al. 2006, Vodicka et al. 2004). Other experimental investigations report associations between one or both variant XPD polymorphisms and reduced DNA repair capacity, especially with respect to damage from UV radiation or bulky chemical carcinogens (Affatato et al. 2004, Au et al. 2003, Qiao et al. 2002 a, b, Spitz et al. 2001, Lunn et al. 2000, Manuguerra et al. 2006). An observational molecular epidemiologic study of genetic variation in XPD reported null associations with polyphenol-DNA adducts and sister chromatid exchange levels (Duell et al. 2000). Other epidemiologic studies found positive associations with chromosomal aberrations, single-strand breaks (Sram et al. 2007, Vodicka et al. 2004), and somatic lung tumor p53 mutation status (Gao et al. 2003, Hou et al. 2003, Mechanic et al. 2005). A recent investigation reported positive associations between XPD polymorphisms and p53 mutations in lung tumors only when assessed in combination with a BER polymorphism (in XRCCI; Gao et al. 2006). Other epidemiological studies report null associations with somatic p53 mutations in bladder tumors (Gao et al. 2010, Ryk et al. 2006, Stern et al. 2006).

Studies report the following results with respect to XPD, PAH-related exposures, and DNA repair endpoints. Experimental investigations have found associations between variant XPD Asp312Asn and Lys751Gln genotypes and "lower DNA repair of plasmid treated with benzo[a]pyrene diol epoxide" (Benhamou and Sarasin 2005) (Spitz et al. 2001) as well as with BPDE-induced DNA adducts in human lymphocytes (Zhao et al. 2008). Null results in the
experimental literature were obtained for associations between Lys751Gln and BPDE-induced chromatid breaks (Lin et al. 2007) and between Asp312Asn or Lys751Gln and DNA repair capacity measured by removal of induced BPDE-DNA adducts (Shen et al. 2006). An epidemiological study reported that variant polymorphisms for Asp312Asn and Lys751Gln reduced XPD mRNA levels in circulating lymphocytes; this association was stronger among smokers, especially for those with the greatest smoking duration and intensity (Pabalan et al. 2010, Wolfe et al. 2007). Another epidemiologic study found associations between variation in XPD and chromosomal aberrations in a population exposed to high levels of environmental air pollution, including carcinogenic PAHs (Sram et al. 2007). Some epidemiologic studies have reported that variants in XPD modify chromosomal and DNA damage levels in coke-oven workers (Cheng et al. 2007, Leng et al. 2004), though another investigation among coke-oven workers found null associations between polymorphisms in XPD and DNA damage endpoints (Wang et al. 2010). Finally, although a few studies report null associations between XPD polymorphisms and PAH-related adducts (Pavanello et al. 2005, Zienolddiny et al. 2006), most of the epidemiologic literature indicates that polymorphisms in XPD are associated with PAH-related DNA adducts in occupational and non-occupational settings (Binkova et al. 2007, Hou et al. 2002, Hu et al. 2008; Matullo et al. 2001, 2003, Palli et al. 2001, Pastorelli et al. 2002, Tang et al. 2002), including in breast tissue (Tang et al. 2002).

**Selected NER Genes and Breast Cancer Risk**

Individual epidemiologic investigations have reported positive associations between breast cancer and polymorphisms in NER genes, including XPD, XPF, ERCC1 and XPG (Crew et al. 2007, Han et al. 2009, Kumar et al. 2003, Lee et al. 2005, Milne et al. 2006, Smith et al. 2003, 2008). Of these genes, null associations have also been reported for variants in XPD, XPF
and XGF (Crew et al. 2007, Jorgensen et al. 2007, Mechanic et al. 2006, Pabalan et al. 2010). A pooled study of 30,000 cases and 30,000 controls of primarily European descent assessed a SNP in XPF (rs744154, not the variant genotyped for this dissertation) and found a null association with invasive breast cancer (Gaudet et al. 2009). Only one investigation examined associations between XPA and breast cancer, reporting a null result (Crew et al. 2007). Two studies assessed relations between variants in ERCC1 and breast cancer, both finding positive associations at least in participant subgroups (Crew et al. 2007, Lee et al. 2005). Several meta-analyses report that SNPs in XPD (Asp312Asn and/or Lys751Gln) are associated with breast cancer risk overall (Manuguerra et al. 2006, Qiu et al. 2010) or among certain ethnic groups (Jiang et al. 2010).

In summary, the literature regarding the selected NER variants and overall breast cancer risk is inconsistent and sparse. Inconsistencies across investigations may be a reflection of the low statistical power common to many genetic epidemiology studies (Pabalan et al. 2010, Zhang et al. 2006). As previously discussed, other methodological issues, study population differences and pathway complexity may underlie inconsistent results as well.

Few studies have examined associations between NER variants and breast cancer in light of PAH-related exposures. The LIBCSP reported elevated ORs for breast cancer among women with at least one variant allele for the XPD Asp\textsuperscript{312}Asn (OR = 1.25, 95% CI: 1.04-1.50) or Lys\textsuperscript{751}Gln polymorphisms (OR = 1.21, 95% CI, 1.01-1.44) (combined effect: OR = 1.33, 95% CI: 1.08-1.64) (Crew et al. 2007, Terry et al. 2004). Associations between variant allele homozygosity and breast cancer were limited to participants with PAH-DNA adduct levels above the median (OR = 1.6, 95% CI: 1.0, 2.6) and current smokers (OR = 2.0, 95% CI: 1.0, 3.8) for the Lys\textsuperscript{751}Gln polymorphism (Terry et al. 2004), and to participants with detectable PAH-DNA adducts for the Asp\textsuperscript{312}Asn polymorphism (OR = 1.8, 95% CI: 1.2, 2.8) (Crew et al. 2007). These
results are consistent with the aforementioned recent meta-analysis which found associations between Lys$^{751}$Gln and Asp$^{312}$Asn polymorphisms and breast cancer only among women with higher levels of aromatic adducts (Pabalan et al. 2010). Results are also consistent with another study population reporting that among participants with certain XPD polymorphisms, PAH-DNA adduct levels were higher in breast tumor tissue than in breast tissue from control participants with benign breast disease (Tang et al. 2002).

The LIBCSP also reported increased breast cancer odds among women with detectable PAH-DNA adducts who were homozygous for the variant allele at the 8092C/A polymorphism in ERCC1 (OR = 1.9, 95% CI: 1.1, 3.3), though no main effect of this variant on breast cancer odds was detected (Crew et al. 2007). Null results were found for polymorphisms in XPA, XPF, and XPG, both with respect to overall breast cancer occurrence and within subgroups of PAH-related exposure or adduct levels (Crew et al. 2007). A large investigation conducted among Caucasians and African Americans, the Carolina Breast Cancer Study, found the strongest associations between cigarette smoking and breast cancer among participants with "four or more 'at risk' genotypes in NER genes;" polymorphisms were genotyped in XPD, XPG, XPF, RAD23B, XPC, and ERCC6 (Mechanic et al. 2006). These interaction results were observed in African Americans only (Mechanic et al. 2006).

**Base Excision Repair (BER)**

**Overview of the BER Pathway**

The BER pathway repairs DNA damage that does not alter the shape of the DNA helix, including from ionizing radiation, deamination, alkylating and methylating agents, ‘non-bulky’ chemical adducts, and oxidative stress (Gammon and Santella 2008, Goode et al. 2002, Robertson et al. 2009, Zhang et al. 2006). It is the main repair pathway correcting oxidative
damage to DNA bases, and may play an important role in the repair of PAH-induced DNA damage (Braithwaite et al. 1998, Gammon and Santella 2008). BER tends to proceed more quickly than NER (Braithwaite et al. 1998). This repair pathway is divided into two subpathways: short-patch BER and long-patch BER (Robertson et al. 2009). The short-patch pathway replaces a single nucleotide, whereas the long-patch pathway replaces at least two nucleotides (Robertson et al. 2009). It is not fully understood what factors lead to short-patch rather than long-patch BER (Robertson et al. 2009). Possibilities include local ATP concentrations as well as lesion properties (Robertson et al. 2009).

In the first step of short-patch BER, a DNA glycosylase recognizes and removes the damaged base, leaving an apurinic/apyrimidinic (AP) site (Braithwaite et al. 1998, Goode et al. 2002, Robertson et al. 2009). DNA glycosylases are relatively specialized (Barnes and Lindahl 2004). For example, oxidative stress-induced 8-oxodeoxyguanosine lesions are excised exclusively by 8-oxoguanine DNA glycosylase, the enzyme encoded by OGG1 (Botieux and Radicella 1999). After DNA glycosylases initiate DNA repair, an AP endonuclease (APEX1) or a bifunctional AP lyase such as OGG1 (which, as mentioned above, is also a DNA glycosylase) cleaves the DNA backbone and removes the damaged sugar residue (Boiteux et al. 1987, Braithwaite et al. 1998, Goode et al. 2002, O'Connor and Laval 1989, Robertson et al. 2009). A DNA polymerase, usually DNA polymerase β, restores the abasic site by synthesizing a new nucleotide using the undamaged DNA strand as a template (Matsumoto and Kim 1995, Robertson et al. 2009). DNA ligation with DNA ligase III completes the short-patch BER process (Kubota et al. 1996, Robertson et al. 2009, Wei et al. 1995).

Although XRCC1 is not an enzyme, it plays a critical coordinating role in short-patch BER by functioning as a scaffolding protein for key enzymes including OGG1, DNA polymerase

The initial long-patch BER steps are similar to those described above: damage recognition and removal by a DNA glycosylase and cleavage of the DNA backbone by an AP endonuclease or an AP lyase (Robertson et al. 2009). In long-patch BER, the latter step recruits DNA polymerase β, δ or ε, proliferating cell nuclear antigen (PCNA), flap structure-specific endonuclease 1 (FEN1) and DNA ligase I (LIG1) (Frosina et al. 1996, Matsumoto et al. 1994, Robertson et al. 2009). The DNA polymerase synthesizes at least two contiguous nucleotides (Robertson et al. 2009). FEN1 then processes the area to make it compatible with DNA ligation by LIG1 (Robertson et al. 2009). PCNA plays a necessary coordinating role in long-patch BER by interacting with both DNA polymerase and FEN1 (Frosina et al. 1996, Klungland and Lindahl 1997, Robertson et al. 2009).

**Selected BER Genes and DNA Repair Phenotype**

Selected BER polymorphisms of interest are those genotyped in the LIBCSP: XRCC1 (Arg194Trp and Arg399Gln) and OGG1 (Ser326Cys). All three of these polymorphisms alter the resulting protein's amino acid sequence.

The XRCC1 Arg399Gln polymorphism, one of the most common and widely studied polymorphisms in XRCC1, was related to PAH-related DNA adduct levels in three epidemiological investigations (Matullo et al. 2001, 2003, Zienolddiny et al. 2006). In one epidemiologic study, the association between elevated bulky adduct levels and the variant Arg399Gln polymorphism, while elevated, did not rise to statistical significance (Palli et al. 2005a).
2001). This polymorphism is also associated with aflatoxin B1-DNA adducts and glycoporphin A somatic mutations (Lunn et al. 1999), polyphenol-DNA adducts, sister chromatid exchanges (Duell et al. 2000) and oral tumor p53 mutation status (Hsieh et al. 2003) in epidemiologic research. It was not associated with glycoporphin A somatic mutations among newborns (Relton et al. 2004), chromosomal aberrations among policemen exposed to relatively high levels of air pollution (Sram et al. 2007), or bladder tumor p53 mutations (Gao et al. 2010, Ryk et al. 2006, Stern et al. 2006). Studies of lung tumor p53 mutations obtained conflicting results (Gao et al. 2003, 2006, Hou et al. 2003). In addition, the Arg399Gln variant genotype was not associated with DNA damage measured by the Comet assay (Leng et al. 2004) or with chromosomal damage measured by micronucleus or micronucleated cell frequency (Leng et al. 2005) among coke-oven workers.

Among in vitro experimental investigations using human lymphocytes, the XRCC1 Arg399Gln variant polymorphism was inversely related to capacity to repair bleomycin-induced DNA damage (Cheng et al. 2009) and was positively associated with sister chromatid exchanges following DNA damage induction by a "tobacco-specific nitrosamine" (Abdel-Rahman and El-Zein 2000). The variant polymorphism was also positively related to etheno-DNA adducts (Li et al. 2006), X-ray induced chromosome aberrations (Au et al. 2003), mutagen (bleomycin and BPDE) sensitivity measured by chromosomal breaks per cell (Wang et al. 2003), and mitotic delay among women with a family history of breast cancer (Hu et al. 2001). DNA repair capacity and chromosome aberrations following UV-induced DNA damage were not related to the Arg399Gln polymorphism in two investigations (Au et al. 2003, Qiao et al. 2002b). However, repair of UV-induced radiation is more relevant to the NER pathway (Braithwaite et al. 1998, Goode et al. 2002).
The *XRCC1* Trp\(^{194}\)Arg polymorphism was not associated with aflatoxin B\(_1\)-DNA adducts or glycoporphin A somatic mutations in an epidemiologic investigation (Lunn et al. 1999). Likewise, in experimental studies, the Trp\(^{194}\)Arg polymorphism was not associated with sister chromatid exchanges (Abdel-Rahman and El-Zein 2000), chromosome aberrations (Au et al. 2003) or mitotic delay (Hu et al. 2001). Another investigation reported that the variant Trp\(^{194}\)Arg allele was inversely associated with mutagen sensitivity measured by chromosomal breaks per cell (Wang et al. 2003). Among coke-oven workers, the SNP was not related to olive tail moment (Leng et al. 2004) and was marginally associated with total micronucleus (\(P = 0.053\)) and micronucleated cell (\(P = 0.050\)) frequencies (Leng et al. 2005).

The most common polymorphism in *OGG1*, Ser326Cys, is associated with reduced DNA glycosylase activity in most (Bravard et al. 2009, Dherin et al. 1999, Hill and Evans 2006, Kohno et al. 1998) but not all (Janssen et al. 2001) relevant investigations. In both *in vitro* and *in vivo* experimental studies (conducted among animals, yeast and bacteria), absence of the OGG1 protein consistently leads to increased rates of somatic GC-TA transversions (Klungland et al. 1999, Michaels and Miller 1992, Thomas et al. 1997, Trapp et al. 2007). Increased tumor formation is observed when both OGG1 and MYH are missing (Trapp et al. 2007, Xie et al. 2004). An epidemiologic study reported that the variant Ser326Cys polymorphism was associated with increased 8-oxodeoxyguanosine DNA lesions (Tarng et al. 2001). Another epidemiologic study did not detect an association between genetic variation in *OGG1* and chromosomal aberrations among policemen exposed to high levels of air pollution (Sram et al. 2007). A recent *in vitro* laboratory investigation reported that the 326Cys polymorphism was associated with increased micronuclei levels and reduced repair of induced oxidative DNA damage (Bravard et al. 2009). Another *in vitro* study also found associations between the

**Selected BER Genes and Breast Cancer Risk**


Few studies have examined associations between BER variants and breast cancer in light of PAH-related exposures. The LIBCSP reported positive associations between the *XRCC1*-399Gln allele and breast cancer among women with detectable PAH-DNA adducts (OR: 1.33, 95% CI: 0.98, 1.84; p for interaction = 0.71) and, unexpectedly, among never smokers (OR = 1.31, 95% CI: 1.01, 1.69; p for interaction = 0.03) (Shen et al. 2005a). The association between the 399Gln allele and breast cancer was strongest among never smokers with detectable PAH-DNA adducts (OR = 1.92, 95% CI: 1.21, 3.07) (Shen et al. 2005a). Null associations were observed between *XRCC1* Arg399Gln or Arg194Trp polymorphisms and overall breast cancer risk, and the study found no evidence of interactions between Arg194Trp and either smoking or PAH-DNA adducts (Shen et al. 2005a). The LIBCSP also reported null results for variants in *OGG1* and breast cancer, which did not vary by active smoking status (Rossner et al. 2006).
ADDRESSING LIMITATIONS OF PREVIOUS INVESTIGATIONS

The current investigation examines the relationship between residential exposure to traffic PAHs and breast cancer, and investigates the impact of genetic variation in DNA repair pathways on this potential association. Previous studies that evaluated relationships between PAH sources and breast cancer have mainly used inconsistently defined surrogates for PAH exposure such as cigarette smoking or meat consumption (Gammon and Santella 2008). These surrogates contain non-PAH carcinogens, are classified and measured using strategies that vary across investigations, and are not consistently linked to breast cancer risk (Gammon and Santella 2008). PAH-DNA adducts are better characterized as biomarkers of PAH exposure. Adduct studies are sparse, but consistently show an association with breast cancer (Gammon et al. 2004a, Gammon and Santella 2008, Rundle et al. 2000a). However, PAH-DNA adducts represent relatively short-term exposures (Gammon et al. 2002c, Rundle et al, 2000a), whereas long-term dose may be more relevant to carcinogenesis (Clark et al. 1997). No known biomarkers represent long-term historical PAH exposure.

Several studies have examined associations between individualized air pollution or traffic emissions exposure estimates and breast cancer, and all found evidence of positive associations (Bonner et al. 2005, Crouse et al. 2010, Nie et al. 2007, Raaschou-Nielsen et al. 2011). One study assessed residential exposure to TSP based on data from a small number of environmental monitoring sites (at birth, menarche, first birth, and 10 and 20 years prior to the case-control interview, as well as cumulative exposure based on data from up to 5 of the aforementioned years) as a rough proxy for ambient PAH exposure (Bonner et al. 2005). The TSP measure was heavily influenced by industrial sources, and is therefore likely to be a less relevant and generalizable PAH surrogate for the general population than residential traffic. Another study
examined associations between residential exposure to traffic PAH emissions (at menarche, first birth, and 10 and 20 years prior to the case-control interview) and breast cancer, using an adapted version of a model developed for the LIBCSP (Beyea et al. 2006, Nie et al. 2007). This geographic model was validated and calibrated against copious field data from the Long Island study (Beyea et al. 2006), and is thus best suited for analyses in this population. A third investigation examined associations between breast cancer and traffic-related nitrogen dioxide emissions in the years 2006, 1996, 1985, and after averaging emissions for 1985 and 1996 (Crouse et al. 2010). The fourth study evaluated long-term exposure to nitrogen oxides in relation to breast cancer risk (Raaschou-Nielsen et al. 2011). Nitrogen oxides are gaseous pollutants, and therefore do not reflect depletion and long-term transport phenomena relevant to carcinogenic particulate traffic air pollution (Beyea et al. 2006, IARC 2010).

Longer-term cumulative measures of air pollution exposure may be more relevant to breast carcinogenesis (Clark et al. 1997). This is especially true because residential ambient PAH and other pollutant exposure levels can vary greatly for a given woman from year to year due to, for example, changes in emissions or moving to a different residence (Beyea et al. 2005).

Furthermore, most previous epidemiological studies of traffic or air pollution have utilized relatively crude exposure surrogates such as distance to the nearest major road, often based on self-reported information, or have used data from sparse environmental monitors to assign exposure levels (Boothe and Shendell 2008, Heinrich et al. 2005, Nie et al. 2007, Savitz and Feingold 1989). Other than the historical geographic model developed for the LIBCSP and adapted for the Western New York Exposures and Breast Cancer study (Nie et al. 2007), no previously reported traffic exposure estimation methods incorporate "excess emissions at intersections and during engine warm-up, which may...be important sources of variation in
exposure" (Beyea et al. 2006) (Carr et al. 2002, Gauderman et al. 2005). It should be noted that cold-engine emissions were dropped from this investigation's exposure model following validation exercises (Beyea et al. 2006). Many investigations have not validated their pollutant exposure estimates.

In contrast, the current investigation examined up to 35 years of residential traffic benzo[a]pyrene exposure estimates. Traffic pollution is a ubiquitous and generalizable PAH exposure surrogate. The current investigation's geographic model of residential traffic PAH exposure was validated against residential soil PAHs and PAH-DNA adduct levels in circulating mononuclear cells among LIBCSP participants, and against local monitored levels of the traffic-related pollutant carbon monoxide (Beyea et al. 2006). The geographic model for this investigation takes into account historical tailpipe PAH emissions, local transportation patterns, meteorological conditions, pollutant dispersion and decay data, cruise and cold-engine emissions, and proximity to intersections (Beyea 1999, Beyea et al. 2006), and allows for calculation of individualized rather than aggregated exposure estimates (Beyea et al. 2006).

Research indicates that modeled traffic exposure estimates predict pollutant concentrations more effectively than cruder traffic exposure proxies or ambient monitoring data (Brauer et al. 2003, Heinrich et al. 2005). Thus, this traffic model likely reduces the exposure misclassification that commonly arises in environmental epidemiology studies (Beyea 1999).

Previous studies suggest that associations between PAHs and breast cancer may be observed only among certain genetic subgroups (Crew et al. 2007, Shen et al. 2005, Terry et al. 2004). However, only one study has examined interactions between a genetic polymorphism (in GSTM1, a metabolic gene) and traffic or air pollution with respect to breast cancer risk (Nie et al. 2005). The current study is thus the first to assess whether genetic variation in DNA repair
pathways impacts the association between traffic or air pollution and breast cancer. In addition, it is the first study to examine effect modification of this potential association by fruit or vegetable intake, or to evaluate associations between traffic or air pollution and breast cancer considered as subgroups defined tumor \( p53 \) mutation status or hormone receptor status.

SUMMARY AND CONCLUSIONS

Breast cancer is the most common non-skin cancer malignancy among US women (ACS 2010). Many risk factors have been established for this disease (ACS 2010, Hankinson et al. 2004). However, other genetic and environmental or lifestyle factors contributing to breast cancer risk have yet to be elucidated. Previous investigations suggest a role for the environment in general, and for traffic and PAHs specifically, in breast carcinogenesis (Beyea et al. 2006, Bonner et al. 2005, Crouse et al. 2010, Gammon et al. 2004a, Laden and Hunter 1998, Lewis-Michl et al. 1996, Li et al. 1996, 2002, Nie et al. 2007, Perera et al. 1995, Rundle et al. 2000a). For example, PAH-DNA adducts (Gammon et al. 2004a, Li et al. 1996, 2002, Perera et al. 1995, Rundle et al. 2000a) and traffic or air pollution exposure variables (Bonner et al. 2005, Crouse et al. 2010, Lewis-Michl et al. 1996, Nie et al. 2007) are consistently associated with breast cancer in epidemiologic studies. However, relevant research is sparse. PAHs are known human carcinogens (for example, in the lung) (IARC 2010) and cause mammary tumors in rodents (el-Bayoumy et al. 1995, IARC 2010). Vehicular traffic is a major source of ambient PAH, especially in or near cities (Bostrom et al. 2002, IARC 1989).

PAH metabolites are directly genotoxic, forming PAH-DNA adducts (Harvey et al. 1996). PAHs can also indirectly cause DNA damage through induction of oxidative stress (IARC 2010). PAH-related DNA damage is corrected through nucleotide excision repair and base excision repair DNA repair pathways (Braithwaite et al, 1998, Goode et al. 2002). If DNA repair
mechanisms fail or are insufficient, the resulting uncorrected DNA damage can contribute to
carcinogenesis through formation of somatic mutations in tumor suppressor genes or proto-
pathways are standard candidate cancer susceptibility genes (Gammon and Santella 2008) and
are hypothesized to modify the effects of environmental exposures on cancer risk (Berwick and
Vineis 2000). Previous research suggests that variants in NER and BER DNA repair genes may
modify relations between PAH-related variables and breast cancer (Crew et al. 2007, Mechanic

The current analysis evaluates the association between vehicular traffic PAHs and breast
cancer, and assesses interactions between traffic PAHs and DNA repair polymorphisms with
respect to breast cancer risk. Findings of an association between traffic PAHs and breast cancer
may have important public health implications given the frequency of breast cancer and the
ubiquity of traffic and PAH exposure across the general population. Currently, no governmental
standards exist regarding ambient environmental PAH levels. Researching the impact of genetic
variation on the association between breast cancer and traffic PAHs has implications with respect
to elucidating possible biological mechanisms underlying this association and distinguishing a
subgroup of women in the general population who may be more susceptible to the carcinogenic
effects of traffic pollution on the breast.
CHAPTER II: RESEARCH METHODS

RESEARCH AIMS, HYPOTHESES, AND STUDY RATIONALE

Breast cancer is the most common malignancy and the second-leading cancer-related cause of death among women in the United States (American Cancer Society 2010). Polycyclic aromatic hydrocarbons (PAHs) are ubiquitous environmental pollutants (Bostrom et al. 2002, IARC 2010). These chemicals have genotoxic and estrogenic properties (IARC 2010, Santodonato 1997), are confirmed human carcinogens (for example, in the lung) (IARC 2010) and cause mammary tumors in laboratory animals (el-Bayoumy et al. 1995, IARC 2010). However, the association between PAHs and breast cancer in humans is unclear (Gammon and Santella 2008). PAH-DNA adducts, a short-term marker of genetic damage from PAHs, are consistently linked with breast cancer in epidemiologic studies (Gammon et al. 2004b, Li et al. 1996, 2002, Perera et al. 1995, Rundle et al. 2000a). Traffic is a major source of PAH exposure, especially in or near cities (IARC 1989, Lioy and Greenberg 1990).

Research regarding traffic or air pollution and breast cancer risk has been sparse, but consistently shows evidence of a positive association (Bonner et al. 2005, Crouse et al. 2010, Lewis-Michel et al. 1996, Nie et al. 2007, Raaschou-Nielsen et al. 2011, Wei et al. 2012). Most research has focused on relatively short-term traffic or air pollution exposure estimates, although breast cancer generally develops over many years (Clark et al. 1997). Since PAHs are genotoxic and can also indirectly induce oxidative DNA damage (IARC 2010), polymorphisms in DNA repair genes may modify associations between traffic PAHs and breast cancer.

This dissertation analysis evaluates the association between estimates of traffic PAH
exposure from a historical geographic model and breast cancer, as well as interactions between traffic PAHs and DNA repair variants, using the resources of a large population-based case-control study: the Long Island Breast Cancer Study Project (LIBCSP) (Gammon et al. 2002b). Clarifying the association between traffic PAHs and breast cancer has important public health implications due to the relatively high incidence of breast cancer and ubiquitous exposure to traffic and PAH pollution in the US general population. No environmental standards for ambient PAH levels have been established in the US.

This dissertation analysis could also increase understanding of breast cancer etiology with respect to DNA repair pathways. Findings of interactions between traffic PAHs and DNA repair polymorphisms would help clarify underlying mechanisms potentially linking PAHs and traffic pollution to breast cancer. Identifying subgroups of women who are especially susceptible to carcinogenic effects of traffic PAH exposure based on genetic characteristics could aid in strengthening the biological plausibility of this relationship and informing public health recommendations for lowering breast cancer risk in vulnerable populations.

**Primary Aim 1: Traffic PAHs and Breast Cancer**

**Study Aim**

Examine associations between breast cancer and estimates of residential exposure to PAHs from vehicular traffic.

**Hypothesis**

Residential traffic PAH exposure estimates are positively associated with breast cancer.

**Rationale**

PAHs are confirmed carcinogens in humans (e.g. in the lung) (IARC 2010, Straif et al.)
2005) and are potent mammary carcinogens in laboratory rodents (el-Bayoumy et al. 1995). PAHs also exhibit estrogenic, pro-oxidant and genotoxic properties (DeMarini 2004, Santodonato 1997, Sorensen et al. 2003). When PAH exposure levels are high or detoxification is insufficient, PAHs bind to DNA forming PAH-DNA adducts, including in breast tissue (Gammon and Santella 2008, Santella 1999). PAHs can also induce oxidative DNA damage (Farmer et al. 2003) and may act as tumor promoters via estrogenic growth stimulation of breast cells (Santodonato 1997). Uncorrected PAH-induced DNA damage can lead to somatic mutations in cancer-related genes, which could in turn contribute to genetic instability and carcinogenesis (Gammon and Santella 2008). PAH-DNA adducts are consistently associated with breast cancer in epidemiologic research (Gammon et al. 2004a, Li et al. 1996, 2002, Perera et al. 2005, Rundle et al. 2000a).

Traffic is an important source of PAH exposure in the general population, especially in or near urban areas (IARC 1989, Lioy and Greenberg 1990). All studies that examined associations between breast cancer and variables related to environmental exposure to ambient PAH or traffic pollution found evidence of positive associations (Bonner et al. 2005, Crouse et al. 2010, Lewis-Michel 1996, Nie et al. 2007, Raaschou-Nielsen et al. 2011, Wei et al. 2012). Although these results are encouraging, each of the four studies had certain weaknesses that the current investigation addresses at least in part. One of the previous investigations calculated residential traffic density using a geographic information system divided into 5 km² grids as their main exposure of interest, and therefore did not examine truly individualized exposure estimates (Lewis-Michl et al. 1996). Another study assessed residential exposure to TSP for up to five individual years as a rough proxy for ambient PAH exposure, using data from a small number of environmental monitoring sites (Bonner et al. 2005). The TSP measure was heavily influenced
by industrial sources, and is therefore likely to be a less relevant and generalizable PAH surrogate for the general population than residential traffic. Another study examined associations between one year intervals of residential exposure to traffic PAHs and breast cancer, using an adapted version of the model developed for the LIBCSP (Beyea et al. 2006, Nie et al. 2007). This geographic model was validated and calibrated against copious field data from the Long Island study (Beyea et al. 2006), and is thus best suited for analyses in this population. Finally, two other investigations examined associations between breast cancer and traffic-related nitrogen oxides (Crouse et al. 2010, Raaschou-Nielsen et al. 2011). Nitrogen oxides are gaseous pollutants, and do not reflect depletion and long-term transport phenomena relevant to carcinogenic particulate traffic air pollution (Beyea et al. 2006, IARC 2010). In contrast to previous studies, this dissertation analysis aims to examine the association between breast cancer and traffic PAHs using exposure estimates from an individualized, long-term (up to 35 years), comprehensive and validated historical geographic model (Beyea et al. 2006).

Primary Aim 2: Effect Modification by DNA Repair Polymorphisms

Study Aim

Examine whether the association between residential traffic PAH exposure and breast cancer is modified by single nucleotide polymorphisms (SNPs) in base excision repair (BER; XRCC1, OGG1) and nucleotide excision repair (NER; XPA, XPD, XPF, XPG, ERCC1) pathways.

Hypothesis

Associations between breast cancer and residential exposure to traffic PAHs are stronger among women with genetic polymorphisms related to reduced DNA repair capacity.
Rationale

PAH exposure can lead to DNA damage through formation of PAH-DNA adducts (Braithwaite et al. 1998) and through induction of oxidative stress (Goode et al. 2002). Since uncorrected DNA damage can lead to somatic mutations in tumor suppressor genes and proto-oncogenes, which could in turn contribute to genetic instability and carcinogenesis, polymorphisms that affect DNA repair capacity may modify associations between PAHs and breast cancer risk (Gammon and Santella 2008). Several classes of biochemical pathways repair DNA damage. The NER pathway generally repairs DNA damage from bulky chemical adducts, while the BER pathway generally repairs oxidative DNA damage (Goode et al. 2002, Braithwaite et al. 1998, Robertson et al. 2009). Hence, both DNA repair pathways could be relevant to PAH-induced carcinogenesis. The LIBCSP investigation is the first to evaluate whether genetic variation in DNA repair pathways modifies the association between traffic or air pollution exposure and breast cancer risk.

Secondary Aim 1: Effect Modification by Fruit and Vegetable Intake and Menopausal Status

Study Aim

Evaluate whether the association between traffic PAHs and breast cancer is modified by menopausal status and by fruit and vegetable intake.

Hypothesis

The association between traffic PAH exposure and breast cancer varies by menopausal status. The association is stronger among women with low fruit and vegetable intake.

Rationale

The risk factor profile for breast cancer differs by menopausal status (Barlow et al. 2006). Studies of traffic or air pollution exposures and breast cancer risk that stratified by
menopausal status report differences in associations for postmenopausal and premenopausal subgroups (Bonner et al. 2005, Nie et al. 2007). One study found a positive association between TSP exposure and breast cancer among postmenopausal, but not premenopausal, women (Bonner et al. 2005). Another study reported associations between traffic exposure at the age of first birth and postmenopausal, but not premenopausal, breast cancer (Nie et al. 2007). Similarly, an investigation of breast cancer and grilled and smoked meat intake, an important PAH source, reported that associations were limited to postmenopausal women (Steck et al. 2007). However, other analyses report associations between traffic emissions or PAH-related occupational exposures and premenopausal breast cancer (Nie et al. 2007, Petralia et al. 1999), and a population-based investigation found a slightly stronger relationship between PAH-DNA adducts and premenopausal versus postmenopausal breast cancer (Gammon et al. 2004a). Thus, it is difficult to predict the direction of any potential variation in the association between traffic PAHs and breast cancer by menopausal status. Possible explanations for differential findings in the literature include etiologic differences between premenopausal and postmenopausal breast cancers, shorter induction time among premenopausal cases, lower traffic or air pollution exposure levels among younger women (for example, due to the introduction of catalytic converters or changing industrial emissions standards), and sample size differences between menopausal status subgroups (Bonner et al. 2005, Nie et al. 2007).

Fruit and vegetables have antioxidant and other chemopreventive properties (for example, related to cell cycle control) that could ameliorate PAH-induced DNA damage and reduce the likelihood of PAH-related carcinogenesis (Amin et al. 2009). Fruit and vegetable intake is linked with decreased breast cancer risk in some, but not all, investigations (Gandini
et al. 2000, Gaudet et al. 2004, Howe et al. 1990, Smith-Warner et al. 2000). A LIBCSP study reported relations between grilled and smoked meat intake and breast cancer that were stronger among women with low fruit and vegetable intake (Steck et al. 2007). Fruit and vegetable consumption was also associated with lower PAH-related DNA adduct levels in several studies, including in the LIBCSP (Palli et al. 2000, Peluso et al. 2000, Shantakumar et al. 2005), and fruit and vegetable components decrease carcinogenic effects of PAHs in animals (Conaway et al. 2005, Kocdor et al. 2005).

Secondary Aim 2: Associations with Breast Tumor Subtypes

Study Aim

Evaluate associations between traffic PAHs and breast cancer subtypes classified according to tumor hormone receptor (ER/PR) status and tumor p53 mutation status.

Hypothesis

Associations between traffic PAHs and breast cancer differ when tumors are subdivided according to hormone receptor status (ER/PR) and somatic p53 mutation status subtypes.

Rationale

Breast cancer risk factor profiles differ by tumor hormone receptor status (Althuis et al. 2004, Colditz et al. 2004, Nasca and Pastides 2008). For example, ER-positive tumors generally show stronger associations with reproductive or hormonal risk factors than ER-negative tumors (Althuis et al. 2004, Chen and Colditz 2007, Colditz et al. 2004, Garcia-Closas and Chanock 2008, Nasca and Pastides 2008). Inconsistent risk factor associations in the breast cancer literature may be attributable to differences in hormone receptor subtype distribution between study populations (Chen and Colditz 2007). PAHs have estrogenic and
antiestrogenic properties (Santodonato 1997).

Two occupational studies assessed relations between PAH exposure and breast cancer categorized by tumor hormone receptor status. Upon stratifying by hormone receptor status, both studies found that associations with breast cancer were limited to estrogen receptor-positive tumors (Labreche et al. 2010, Petralia et al. 1999). In the general population, however, research regarding PAH-related exposures and breast tumor hormone receptor status has been inconsistent (Althuis et al. 2004, Gammon et al. 2002c, 2004a,b, Manjer et al. 2001, Morabia et al. 1998, Steck et al. 2007, Terry and Rohan 2002). A 2004 review concluded that, based on the published literature, associations between cigarette smoking and breast cancer do not differ with respect to tumor hormone receptor status (Althuis et al. 2004), though individual investigations report evidence of differential associations (e.g. Britton et al. 2002, Gammon et al. 2004b). The only study to evaluate associations between traffic exposure and breast cancer categorized according to tumor hormone receptor status did not find evidence of differing associations by tumor subtype (Nie et al. 2007).

Lehman and Harris 1994, Mahadevan et al. 2003). It should be noted that p53-mutation positive breast tumors are more likely to be hormone-receptor negative (Olivier et al. 2006).

Cigarette smoking was positively associated with breast tumor p53 mutation status in two studies (Conway et al. 2002, Van Emburgh et al. 2008). However, findings from a larger study suggest that PAH-related variables (active or passive smoking, grilled/smoked meat intake, PAH-DNA adducts) are not associated with p53 mutation-positive breast cancer. In fact, "participants with breast tumor p53 mutations were less likely to be exposed to PAH-related sources than were participants with p53 mutation-negative cancer" (Mordukhovich et al. 2009). Two large general population studies report possible associations between PAH-related exposures and a differing set of specific breast tumor p53 mutation types (Conway et al. 2002, Mordukhovich et al. 2009).

PAHs are an important component of traffic pollution and exhibit estrogenic, antiestrogenic, pro-oxidant and genotoxic properties (Farmer et al. 2003, Jeng et al. 2010, Santodonato 1997). Thus, it is plausible that associations between traffic PAHs and breast cancer vary with respect to tumor hormone receptor status or p53 mutation status. It is difficult to predict the direction of this potential variation due to opposing estrogenic and antiestrogenic effects (Santodonato 1997) as well as the role of genotoxic and pro-oxidant effects (IARC 2010) of PAHs. However, it is likely that the genotoxic and pro-oxidant properties of PAHs are dominant over estrogenic or antiestrogenic properties at the environmental exposure levels in the current investigation (Bostrom et al. 2002).

LONG ISLAND BREAST CANCER STUDY PROJECT (LIBCSP)

The parent study for this dissertation analysis is the Long Island Breast Cancer Study Project (Gammon et al. 2002b). This is a large population-based case-control study initiated
following a federal mandate to investigate environmental risk factors for the high breast cancer rates in Nassau and Suffolk counties on Long Island, New York. Primary aims of this research effort were to evaluate associations between breast cancer and exposure to organochlorines and PAHs (Gammon et al. 2002b). The parent study received IRB approval from all relevant institutions, and all women gave written informed consent prior to study participation (Gammon et al. 2002b). Study participants were primarily White (specifically, 93% White, 5% African American, 2% other; 4% self-identified as Hispanic), consistent with the race distribution in Nassau and Suffolk counties (Gammon et al. 2002b).

LIBCSP study participants were adult, English-speaking women ages 20 to 98 residing in Nassau or Suffolk counties on Long Island, New York (Gammon et al. 2002b). Eligible cases were women with no prior personal history of breast cancer who were diagnosed with primary invasive or in situ breast cancer between August 1, 1996 and July 31, 1997. Cases were recruited using rapid case ascertainment through frequent contact (at least twice per week) with the pathology departments of 31 hospitals in Long Island and New York City. The subjects' physicians were contacted to confirm breast cancer diagnoses and obtain permission to contact the cases. Eligible controls were women with no personal history of breast cancer residing in Nassau or Suffolk counties. They were randomly selected for recruitment to the LIBCSP via random digit dialing (using Waksberg's method; Waksberg 1978) for women under age 65 (the screening response rate was 78%; Gammon et al. 2002b) and via Health Care Finance Administration records for women 65 years or older. Controls were frequency matched to cases based on the expected age distribution among cases. Eligible cases and controls were mailed information about the LIBCSP. Recruiters then contacted the eligible participants to schedule the interview mainly via telephone, though some control participants were contacted in person.
(Gammon et al. 2002b). In total, 1,508 cases (82.1% of eligible case participants) and 1,556 controls (62.7% of eligible control participants) completed the case-control interview. Participation was negatively associated with age for both cases and control participants (Gammon et al. 2002b).

The main study interview was conducted by trained interviewers at the subjects' residences for both cases and controls. On average, these interviews occurred 96 days after diagnosis for breast cancer cases and 167 days after identification for controls (Gammon et al. 2002b). The case-control questionnaire collected detailed information on a wide range of variables (such as known or suspected breast cancer risk factors, as well as potential confounders or effect modifiers), including lifetime residential history in Nassau and Suffolk counties, demographic characteristics, medical history, family history of cancer, reproductive history, lifetime physical activity, cigarette smoking (both active and passive), lifetime alcohol consumption, menopausal status, body size across the life-course, and lifetime intake of grilled and smoked meat (Gammon et al. 2002b). Variables relevant to this dissertation analysis are discussed in more detail in Chapter II. Following the case-control interview, 98.2% of cases and 97.6% of controls completed a self-administered, validated Block food frequency questionnaire (FFQ) (Block et al. 1986, 1990, Potischman et al. 1997), which assessed usual dietary and supplement intake during the year preceding the reference date (date of diagnosis for cases, date of identification for controls) (Gammon et al. 2002b, Shen et al. 2005b).

At the same study visit, LIBCSP interviewers, who were either nurses or certified phlebotomists, collected blood samples for 73% of both case and control participants (Gammon et al. 2002b). These samples were stored and processed as described in Chapter II. Blood samples were collected in order to genotype common variants in candidate breast cancer
susceptibility genes (for example, in DNA repair and xenobiotic metabolism pathways), to evaluate PAH-DNA adducts in circulating mononuclear cells (Gammon et al. 2002c), and to assess other biomarkers of interest (Gammon et al. 2002b).

The interviewers also conducted residential sampling of outdoor soil and indoor carpet dust among a subset of participants who had lived at their current residence for at least 15 years (Gammon et al. 2002b). Soil samples were collected and analyzed for benzo[a]pyrene and other PAH species using high-resolution gas chromatography/mass spectroscopy for 360 cases and 356 control participants (Beyea et al. 2006, Gammon et al. 2002b). Carpet dust samples were collected and analyzed for PAH content using gas chromatography/mass spectroscopy selected ion monitoring for 320 cases and 356 controls (Beyea et al. 2006, Gammon et al. 2002b). This component of the parent study was executed primarily in order to validate the LIBCSP geographic model of traffic PAH exposure (Beyea et al. 2006, Gammon et al. 2002b). Participants who lived at their current residence for 15 or more years differed from other participants with respect to age, race, education, income, parity, menopausal status, age at menopause, lactation history, smoking status, body mass index, alcohol intake, oral contraceptive use, and hormone replacement therapy use (Gammon et al. 2002b).

Most LIBCSP cases (n = 1473, or 97.7% of case participants) signed medical release forms that allowed researchers abstract their medical records in order to confirm study eligibility and ascertain clinical characteristics of their disease, including tumor hormone receptor status (Gammon et al. 2002b). Medical records were successfully abstracted for 1,402 cases (95.2% of case participants). Case participants primarily presented with hormone responsive breast tumors (ER+/PR+: n = 583 [59%]; ER+/PR-: n = 143 [14%]; ER-/PR+: n = 52 [5%]; ER-/PR-: n = 212 [21%]). Archived paraffin-embedded tumor tissue, which was used to analyze tumor markers,
was obtained from participating hospitals for 962 cases who signed medical release forms. For 
p53 analyses, sufficient tumor DNA was available for 859 cases (Rossner et al. 2008). 
Laboratory procedures for ascertaining somatic p53 mutations are described in more detail in 
Chapter II.

Data coding and data entry were completed by Westat, Inc. in Bethesda, MD (Gammon et 
al. 2002b). Quality control measures employed by the LIBCSP include questionnaire data 
verification as well as data range and logic checks, also completed by Westat, Inc. (Bethesda, 
MD) (Gammon et al. 2002b). In addition, a random sample of 20% of participants were 
telephoned following their study visit in order to (1) confirm that they had been interviewed for 
the LIBCSP, (2) confirm the length of their study visit, and (3) briefly re-interview the 
participant via telephone (Gammon et al. 2002b). Identical data collection and storage strategies 
were implemented for cases and control participants in order to minimize issues of differential 
misclassification. Quality control measures for the laboratory and geographic modeling 
components of the LIBCSP are described below.

The current analysis uses existing data from the LIBCSP parent investigation, and 
likewise employs a case-control design. Eligible subjects for the proposed investigation are 
parent study participants with complete information on all variables necessary for any given 
study aim. Appendix Table A2.1, adapted from Gammon et al. 2002b, shows the distribution 
of study participant characteristics by case-control status in the LIBCSP.

RESULTS FROM THE LIBCSP: PAH-RELATED EXPOSURES AND BREAST CANCER

The LIBCSP has previously evaluated associations between several PAH-related 
exposures and breast cancer risk. Briefly, the LIBCSP reports positive associations between 
breast cancer and long-term passive smoking (but not active smoking), grilled or smoked meat
intake, and PAH-DNA adducts (Gammon et al. 2004a, 2004b, Steck et al. 2007). Relevant study results are described in more detail below. Findings from the LIBCSP suggest that PAH-related exposures (smoking, grilled and smoked meat intake, and adducts) may not be associated with overall breast tumor p53 mutation status, though possible associations with specific mutation subgroups were detected (Mordukhovich et al. 2009).

Active and Passive Cigarette Smoking

The LIBCSP reported a positive association between breast cancer and long-term passive smoking (OR = 2.10, 95% CI: 1.47, 3.02 among nonsmokers who lived with a smoking spouse for over 27 years; Gammon et al. 2004b), as well as between a combined measure of active and passive cigarette smoking and ER/PR positive breast cancer (OR = 1.42, 95% CI: 1.00, 2.00; Gammon et al. 2004b). In addition, the LIBCSP reported evidence of interactions between smoking and SNPs in XPD and XRCC1, but not OGG1, with respect to breast cancer risk (Rossner et al. 2006, Shen et al. 2005a, Terry et al. 2004). The LIBCSP found null associations between breast cancer and a variety of other current or former active smoking categorizations, including smoking only prior to first pregnancy, as well as most passive smoking measures (Gammon et al. 2004b).

Grilled and Smoked Meat Intake

The LIBCSP reported positive associations between lifetime total grilled or smoked meat intake and postmenopausal, but not premenopausal, breast cancer (OR = 1.47, 95% CI: 1.12, 1.92 for the highest vs. the lowest tertile of intake; Steck et al. 2007). This relation was stronger among women with low fruit and vegetable consumption (OR = 1.74, 95% CI: 1.20, 2.50 for the highest vs. the lowest tertile of intake of grilled or smoked meat, p-interaction > 0.05), but did not appear to vary with respect to ER/PR status (Steck et al. 2007). Elevated effect estimates
were reported for the intake of red, but not white, grilled or smoked meat (Steck et al. 2007). This study failed to find associations between breast cancer and dietary intake of benzo[a]pyrene in the past year, "with the possible exception of benzo[a]pyrene from meat among postmenopausal women whose tumors were positive for both estrogen receptors and progesterone receptors (OR = 1.47; CI = 0.99-2.19)" (Steck et al. 2007).

**PAH-DNA Adducts**

The LIBCSP reported positive associations between breast cancer and PAH-DNA adducts in circulating mononuclear cells (OR = 1.41, 95% CI: 1.07, 1.86 for the highest vs. the lowest quantile, and OR = 1.29, 95% CI: 1.05, 1.58 for detectable vs. nondetectable adduct levels; Gammon et al. 2004a). This association did not consistently vary by levels of other PAH exposure sources or with respect to hormone receptor status, and was slightly stronger among premenopausal women (OR = 1.56, 95% CI: 1.09, 2.23 for detectable adducts, p for interaction > 0.05) (Gammon et al. 2004a). No dose-response relationship with breast cancer was observed when evaluating quantiles of PAH-DNA adduct levels (Gammon et al. 2004a). The LIBCSP reports evidence of interactions between PAH-DNA adducts and SNPs in *XPD, XRCC1* and *ERCC1*, but not *XPA*, *XPF* or *XPG*, with respect to breast cancer risk (Crew et al. 2007, Shen et al. 2005a, Terry et al. 2004).

The LIBCSP found positive associations between detectable PAH-DNA adducts and current and past smoking as well residential soil PAHs. Null associations were reported for dietary PAHs and modeled ambient benzo[a]pyrene levels, and detectable adducts showed a negative association with PAH levels in residential dust (Shantakumar et al. 2005). Among women with detectable PAH-DNA adducts, PAH-DNA adduct levels were positively associated with traffic benzo[a]pyrene exposure levels derived from the LIBCSP geographic model,
described below (Beyea et al. 2006).

Traffic PAHs

Preliminary analyses in the LIBCSP show a subgroup of 1% of cases with much greater estimated traffic PAH exposure than all control women for absolute cumulative exposure, relative exposure, and maximum annual exposure (years 1960-1990; Dr. Jan Beyea, personal communication 2009).

PAH EXPOSURE ASSESSMENT: A HISTORICAL GEOGRAPHIC MODEL

Brief Overview

A historical model of residential PAH exposure from vehicular traffic (years 1960–1995) was developed for the LIBCSP by Dr. Jan Beyea, a consultant at CiPI (Consulting in the Public Interest) (Beyea et al. 2006). As described in detail below, this geographic model incorporates historical US vehicular PAH emissions data, information on traffic and transportation patterns in the New York metropolitan area, Long Island meteorological variables, and pollutant dispersion factors (Beyea et al. 2006). The model was validated and calibrated against extensive field data from the LIBCSP (Beyea et al. 2006). The LIBCSP includes 2,608 participants with traffic PAH exposure estimates for the year 1995 and 1,719 participants with exposure estimates spanning the years 1980 to 1995.

Data Collection and Geocoding

During the case-control interview, trained interviewers collected information regarding participant lifetime residential history in Nassau and Suffolk counties on Long Island, NY. Specifically, interviewers recorded all addresses at which a participant resided for at least one year in these counties, including prior to 1960 (Beyea et al. 2005). Addresses were then geocoded using BLR software (BLR Data, Inc. Lebanon, NH), and poorly coded addresses were
cleaned manually. A total of 8,321 residential addresses were recorded at the case-control interview (median = 2, maximum = 17), of which 5,501 addresses geocoded successfully to the level of an individual residence. Some recorded addresses were incomplete; 300 recorded addresses did not include a street name and 1,715 included a street name but no address number (Beyea et al. 2005). Of the 6,306 complete addresses, 87% geocoded successfully. Failure to geocode complete addresses is likely due to inaccurate residential address reporting as well as changes in the built environment over time (Dr. Jan Beyea, personal communication 2009). The geocoding success rate was 65% for 1960 addresses and 85% for current addresses, and ranged from 60% to 85% in the period between the year 1960 and the study interview (Dr. Jan Beyea, personal communication 2009). Most LIBCSP participants provided at least one address that geocoded to street level (for addresses between the years 1960 and 1990, n= 2,601; for 1995 addresses, n = 2,655).

Historical tailpipe emission parameters were based on measurements conducted in road tunnels throughout the US (Boston, Baltimore, Los Angeles, Berkeley, and New York City/New Jersey) and checked using individual vehicles run on dynamometer test beds (Beyea et al. 2008). Data used to generate cold-engine (i.e. during engine warm-up) emissions estimates included the "number of households at the census block level and extracts from the travel-diary database of the National Personal Transportation Survey," which yielded information on "the number of cold starts per household per hour per day" (Beyea et al. 2005) (US Department of Transportation [USDOT] 1996). Data on warm-engine (i.e. cruise condition) transportation patterns were derived from more than 13,000 measurements of annual average daily traffic recorded on lists and maps from state and county traffic department records (Beyea et al. 2005).
Model Development

This study evaluated a road network consisting of approximately 500,000 straight-line street segments within the New York metropolitan area, which encompasses 22 counties in New York, New Jersey, and Connecticut (Beyea et al. 2005). In the default, uncalibrated model, estimated benzo[a]pyrene emissions for each road segment (in µg/km per day) were calculated as a product of historical traffic count on the roadway (vehicles per day) and historical average US tailpipe emissions per vehicle under cruise conditions (µg/km). Cruise emissions were calculated for 1960, 1970, 1980, and 1990. Estimates for other years were derived via interpolation or extrapolation (Beyea et al. 2005). Emission estimates were further adjusted for non-cruise conditions (i.e. cold-engine operation and acceleration and deceleration at intersections) using parameters from the literature (Beyea et al. 2006) (e.g. Ahlvik et al. 1997). Emissions were adjusted for intersection traffic using a simplified version of a model originally developed for CO (Beyea et al. 2005, Nelli et al. 1983). Cold-engine operation emissions were estimated with a modeling approach developed by Dr. Beyea (Beyea et al. 2005).

Vehicular PAH emission estimates were converted into predicted residential ambient concentrations of benzo[a]pyrene for each participant (in ng/m³) using standard meteorological dispersion and deposition models (Beyea et al. 2006). Up to a distance of 100 meters from a given road, contributions of estimated emissions from each road segment to surrounding traffic-related benzo[a]pyrene concentrations were calculated using highway line-source models (Chock 1978). The Chock model was used because of its good fit "to tracer concentrations near the Long Island Expressway" (Beyea et al. 2005). Benzo[a]pyrene concentrations farther than 100 meters from a road were calculated using a standard Gaussian plume model with Briggs dispersion parameters (Beyea et al. 2006, Catalano et al. 1987, Viegele and Head 1978), which
incorporated data on variables affecting PAH dispersion such as wind speed and direction, rain washout, photo decay, and particle deposition (Beyea et al. 2006, Fan et al. 1995, Huang et al. 1995, National Council on Radiation Protection and Measurement [NCRP] 1993, Ramsdell et al. 1994). For consistency, the same street maps used for geocoding were used to model dispersion of traffic pollution.

Meteorological data for the dispersion model were collected in Brookhaven National Laboratory in Suffolk County for the year 1993 (Beyea et al. 2005). For some variables, raw meteorological readings were modified for model inclusion using information from the literature (Beyea et al. 2006, Fan et al. 1995, Huang et al. 1995, National Council on Radiation Protection and Measurement [NCRP] 1993, Ramsdell et al. 1994). For example, rainfall measurements were converted into washout rates using a standard function (Ramsdell et al. 1994). Sensitivity analyses indicate that changing the year (1990 vs. 1993) or location (MacArthur Airport vs. Brookhaven National Laboratory) of meteorological data collection does not appreciably alter residential traffic PAH exposure estimates (Beyea et al. 2005).

The LIBCSP model calculated traffic benzo[a]pyrene emissions on major roads within 80 kilometers of Nassau and Suffolk counties (Beyea et al. 2006). In addition, a background term was included to account for more distant roads and other ambient PAH sources (Beyea et al. 2006). Indoor traffic PAH estimates were derived by multiplying outdoor residential exposure estimates "by an average building penetration factor of 0.75" (Beyea et al. 2005), based on values from the literature (Long et al. 2001). The literature informing model parameter values was sparse for some variables such as distance of increased intersection emissions, the ratio of intersection emissions to cruise emissions, and photo decay rate (Beyea et al. 2006). However, all model parameters were calibrated against field data following validation exercises (see
PAH exposures were calculated beginning in 1960, the earliest year for which reliable vehicular PAH emissions and traffic data exist across the US and in the New York metropolitan area, respectively (Beyea et al. 2005). The year 1995, shortly prior to the LIBCSP reference date (1996-1997; date of diagnosis for cases, date of recruitment for controls), was chosen as the end period. PAH exposure measures of interest, including for sensitivity analyses, are discussed in greater detail below (Chapter II).

Model Validation and Calibration

The model was validated by "checking predictions of quantities closely related to PAH air concentration" (Beyea et al. 2005) against relevant field data from the LIBCSP (Beyea et al. 2006). Specifically, the LIBCSP evaluated (1) residential benzo[a]pyrene soil levels among women living at their current address for at least 15 years, (2) PAH-DNA adduct levels in circulating mononuclear cells among women with detectable adducts, (3) US EPA monitoring data for carbon monoxide, a traffic-related pollutant, and (4) residential carpet PAHs among a subset of women living in their current home for at least 15 years (Beyea et al. 2006). Measured ambient PAH concentrations were not available for validation purposes in the study area.

The model accurately predicted both soil PAHs and PAH-DNA adduct levels (Beyea et al. 2006). This was accomplished by modifying the geographic model slightly in order to predict LIBCSP participants' residential soil PAH concentrations and PAH-DNA adduct levels in mononuclear cells, and running regressions of these predictions against field measurements from the LIBCSP (Beyea et al. 2006). Both validation exercises indicated that intersection emissions were important contributors to traffic PAH levels (80% and 40% of the total estimated exposure, respectively) (Beyea et al. 2006). The model also accurately predicted carbon monoxide
concentrations at a local EPA monitor (Beyea et al. 2006).

The model was not successfully validated against the carpet dust data (Beyea et al. 2006). This finding is consistent with research indicating that measured residential ambient PAH concentrations do not correlate with house dust PAH levels (Fromme et al. 2004). Carpet dust PAH level is also the only PAH-related measurement from the LIBCSP that does not decline with decreased urbanization, although overall carpet dust levels do correspond to the urbanization gradient in the study area (Beyea et al. 2006). Instead, benzo[a]pyrene levels in carpet dust peak at an intermediate point along this gradient (Beyea et al. 2006). In addition, carpet benzo[a]pyrene showed a negative association with detectable PAH-DNA adducts in a previous LIBCSP report (Shantakumar et al. 2005). These inconsistent carpet dust data may suggest, for example, “possible spatial confounding with PAH-containing contamination tracked in from outdoors or unmodeled cooking sources” (Beyea et al. 2006).

Default model parameters were chosen based on sometimes sparse or variable information from the literature (Beyea et al. 2006). Therefore, following validation, the model was calibrated against relevant field data through chi-squared minimization (Press et al. 1992), which "was carried out while simultaneously...optimizing a range of other model parameters such as washout rate, particle deposition rates, photo decay rates, and intersection distance" (Beyea et al. 2006). Calibrating the model based on soil concentrations emphasized the contribution of warm-engine emissions, whereas the adduct-calibrated model emphasized cold-engine emissions (Beyea et al. 2006).

The soil-calibrated PAH prediction model provided the best fit to field data ($r^2 = 0.86$ for the soil-calibrated model, $r^2 = 0.58$ for the adduct-calibrated model) (Beyea et al. 2006). Soil-calibrated estimates were therefore be used for regressions against breast cancer risk. These
estimates ignore the contribution of cold-engine emissions. The soil-calibrated model also incorporates a spatially varying background term (as opposed to a constant background term) which is proportional to estimated traffic PAH emissions from counties along metropolitan New York roadways, excluding Nassau, Suffolk, and Queens Counties (Beyea et al. 2006). Other relevant parameters in the soil-calibrated model include a deposition velocity of 0.007 m/s, elevated intersection emissions for a stretch of 12.5 meters, and washout and photo decay rates that were one-half and one-quarter, respectively, of the rates in the default model (Beyea et al. 2006). Intersection emissions account for 80% of average estimated PAH exposure levels in the soil-calibrated model (Beyea et al. 2006).

Treatment of Missing Data

The current investigation evaluates the following residential traffic PAH exposure classifications as the a priori main exposure variables of interest: (1) 1995 exposure and (2) 1980-1995 exposure with up to 20% of residential PAH exposure imputed for a given participant. Sensitivity analyses evaluated additional PAH exposure variables for the years spanning 1980 to 1995 (complete case analysis and up to 30% multiple imputation), 1960 to 1995 (complete case analysis, and up to 20% and 30% multiple imputation), and 1960 to 1990 (complete case analysis, and up to 20% and 30% multiple imputation).

To address missing data, the PAH exposure metric was divided into three components (Dr. Jan Beyea, personal communication 2011, 2012): (1) pre-1960 exposure, (2) exposure after 1960 but before arrival into the study area, and (3) exposure after arrival into the study area, though sometimes exposure could not be calculated during this period because a residential address did not geocode to street level.

Residential traffic PAH exposure from birth to the year 1960 was accumulated in the
period prior to reliable information on traffic patterns or PAH tailpipe emissions. Thus, for pre-
1960 exposure, age at diagnosis or identification was used as a surrogate for time since the start
of the unmeasured traffic PAH exposure (Dr. Jan Beyea, personal communication 2009, 2012,
Greenland 1995).

Another component of participants' PAH exposure spans the period between the year
1960 or birth, whichever occurred later, and the year in which a participant moved into the study
area. Reliable information regarding US tailpipe PAH emissions exist for this period. However,
subjects' PAH exposures could not be calculated during this time because of unknown residence
location. For this second exposure period, the pollutant dispersion function was treated as a
constant, and data on historical US tailpipe PAH emissions (specifically, the integral of the
historical emission curve) were used to develop an exposure surrogate. This exposure surrogate
was calculated separately for cases and controls and was standardized to the average calculated
traffic benzo[a]pyrene concentration on Long Island during the relevant time period (Dr. Jan
Beyea, personal communication 2012). Variance was added during multiple imputation and
matched to the variance in traffic PAH exposure for study participants living on Long Island
during the same time period, based on five-year intervals; variance was not calculated separately
by case-control status (Beyea, personal communication 2012).

Following a participant's arrival into the study area, the LIBCSP historical geographic
model generates individualized PAH exposure estimates at each residence that successfully
geocoded to street level. A participant's cumulative post-arrival exposure was calculated by
summing exposure estimates from all residences through the year 1995. However, because some
residences failed to geocode to street level, as described above, a portion of a given participant's
post-arrival residential traffic PAH exposure may be missing. This missing data issue can be
counteracted in part by using multiple imputation (Dr. Jan Beyea, personal communication 2012, Rubin 1996). Specifically, the transfer function, defined as a function that converts emissions on roadways to exposures at individual residences (Dr. Jan Beyea, personal communication 2012), was imputed for each missing post-arrival address using interpolation or extrapolation from a woman's known residential addresses. When prior and subsequent addresses were available, draws were made from the assigned log-normal variance. When only one address was available, extrapolation was used and a zero variance was assigned (Dr. Jan Beyea, personal communication 2012). Imputation of the transfer function by census place was conducted as well, for the purposes of a sensitivity analysis.

For regression analyses, participants' percentage of imputed exposure, including both prior to and following arrival into the study area, was restricted to 0%, 20% or 30%, depending on the specific analysis. Participants with greater than a pre-specified percentage of their PAH exposure information missing for the time period of interest were dropped from relevant analyses. When using partially imputed variables in regressions, odds ratios and confidence limits were combined over 30 imputations using Rubin's rules (Rubin et al. 1996).

GENOTYPING ASSAYS

Selected SNPs

The polymorphisms evaluated in this investigation are located in genes from BER (X-ray repair cross complementing group 1 [XRCC1] and 8-oxoguanine DNA glycosylase [OGG1]) and NER (xeroderma pigmentosum groups A, D, F, and G [XPA, XPD, XPF, XPG] and excision repair cross-complementing group 1 [ERCC1]) pathways. The SNPs of interest for this dissertation analysis were previously genotyped in the LIBCSP study population: XPA -4A/G (rs1800975), ERCC1 8092C/A (rs3212986), ERCC2/XPD Lys751Gln (rs3212986) and
Asp312Asn (rs1799793), ERCC4/XPF Arg415Gln (rs1800067), ERCC5/XPG Asp1104His (rs17655), XRCC1 Arg194Trp (rs1799782), Arg399Gln (rs25487), and OGG1 Ser326Cys (rs1052133).

Genotyping data availability varied by polymorphism, ranging from 2,050 to 2,177 participants with complete information for a given genetic variant (i.e. 91-97% of women who donated a blood sample; Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005, Terry et al. 2004). Missing genotype data was mainly due to insufficient DNA available for assays (Crew et al. 2007, Terry et al. 2004). All relevant SNPs were previously reported to be in Hardy-Weinberg equilibrium (HWE) among control LIBCSP participants (Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005, Terry et al. 2004).

Some of the selected NER polymorphisms alter the resulting protein's amino acid sequence (XPD Lys751Gln and Asp312Asn, XPF Arg415Gln, and XPG His1104Asp). Others are located outside of coding regions (XPA -4A/G and ERCC1 8092C/A). The LIBCSP has reported modest positive associations between breast cancer and variant alleles in XPD Asp312Asn and Lys751Gln (OR range: 1.21 – 1.25) (Crew et al. 2007, Terry et al. 2004). No associations between polymorphisms in XPA, XPF, ERCC1 or XPG and overall breast cancer risk were reported for the LIBCSP (Crew et al. 2007).

All three selected BER polymorphisms (XRCC1 Arg194Trp, Arg399Gln, OGG1 Ser326Cys) alter the resulting proteins' amino acid sequences. No associations were found between variants in XRCC1 or OGG1 and overall breast cancer risk in the LIBCSP (Rossner et al. 2006, Shen et al. 2005).

**Blood Sample Collection and Storage**

Study interviewers, who were either nurses or certified phlebotomists, obtained a 40 mL
non-fasting blood sample at the study visit from 73.1% of cases (n = 1,102) and 73.3% of control participants (n = 1,141) (Gammon et al. 2002b). Blood donation among LIBCSP participants was positively associated with white race, alcohol intake, hormone replacement therapy use, lactation history, and mammography (Gammon et al. 2002b). Donation was negatively associated with former smoking and age and was unrelated to case-control status (Gammon et al. 2002b).

Blood samples were collected and stored in EDTA-treated, lavender-top tubes at room temperature and shipped overnight to a laboratory in Columbia University (principal investigator: Dr. Regina Santella) (Gammon et al. 2002b,c). There, the blood samples were aliquoted into red blood cells, mononuclear cells, granulocytes, and plasma and were stored in -80°C freezers using preprinted bar code labels with subjects’ randomly assigned study identification numbers. For most participants, blood samples were processed and aliquoted within 24 hours of collection. As a quality control measure, laboratory personnel were unaware of the case-control status of these samples.

**DNA Extraction and Genotyping**

Pelleted mononuclear cells were separated from whole blood by Ficoll (Sigma Chemical Co., St. Louis, MO), washed twice with phosphate-buffered saline (PBS), and frozen at -80°C until DNA isolation (Gammon et al. 2002c). DNA isolation was performed by standard methods: extraction using phenol and chloroform isoamyl alcohol, followed by RNase and proteinase K treatment (Ahn et al. 2004, Gammon et al. 2002c, Shen et al. 2005a). Several high-throughput genotyping methods are available for epidemiological studies. The genotyping methods used in the LIBCSP have changed over time. All methods relevant to the SNPs selected for this dissertation analysis are described in detail below.
Fluorescence Polarization (FP)

Polymorphisms in ERCC1, XRCC1, OGG1 and XPD (Lys751Gln) were genotyped using FP (Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005a, Terry et al. 2004). This method differentiates between alleles in a polymorphic gene "by the template-directed incorporation of a dye-labeled dideoxynucleotide onto an oligonucleotide primer that anneals just 5’ to the polymorphic base" (Shen et al. 2008) (Chen et al. 1999). FP has the advantage over alternative genotyping methods of not requiring specialty probes (Crew et al. 2007).

The following forward and reverse primers were used for polymerase chain reaction (PCR) amplification: ERCC1 8092C/A: forward 5’-TAGTTCTCATGTTTCCCG-3’, reverse 5’-TGAGCCAATTGCAGCCACT-3’; XRCC1 Arg194Trp: forward 5’-ATGAGAGCGCCAACTCTCTG-3’, reverse 5’-CTACCCTCCTCCCTCAGGACC-3’; XRCC1 Arg399Gln: forward 5’-CCCCAAGTGACGAGGCTCAGGTC-3’, reverse 5’-ATTGCCGACAGCGATGAGGATAAG-3’; OGG1 Ser326Cys: forward 5’-TCTCCACCTCCAACACTGTCA-3’, reverse 5’-ATCTAGGCTCCTCAGGCGGCTTTT-3’; XPD Lys751Gln: forward 5’-CCCTCTCCTCCTTTTCTCTGTT-3’, reverse 5’-GGCAAGACTCAGGAGTCAACC-3’.

PCR amplification of genomic DNA was completed using a solution of water, 25 mmol/L MgCl₂, 0.1 μL of HotStar Taq polymerase at a concentration of 5 units/mL (Roche Molecular Biochemicals, Indianapolis, IN), 1 μL of 10× PCR buffer, 25 ng of DNA, 0.2 μL of forward and reverse primers (at a concentration of 8 pmol/μL) and 0.25 μL of deoxynucleotide triphosphates (at a concentration of 10 mmol/L) (Roche Molecular Biochemicals) (Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005a, Terry et al. 2004).

Forward and reverse primers as well as deoxynucleotide triphosphates were processed
using 1 unit of shrimp alkaline phosphatase (at a concentration of 1 unit/μL) (Roche Molecular Biochemicals), 1 unit of \textit{Escherichia coli} exonuclease I (at a concentration of 10 units/μL) (US Biochemical, Cleveland, OH), 1 μL of 10× buffer, and 7.9 μL of water (Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005a, Terry et al. 2004). This step was undertaken in order to "[clean] the PCR products by digesting the excess primers and deoxynucleotides" (Shen et al. 2005a).

Reverse extension primers were 5'-CTACACAGGCTGCTGCTG CTGCT-3', 5'-CGGGGGCTCTCTTCTTCAGC-3', 5'-GTCGGCGGCTGCCCTCCC-3', 5'-CCCTCCTAC AGGTGCTGTTC and 3'-CTGAGCAATCTGCTCTATCCTCT-5' for \textit{ERCC1} 8092C/A, \textit{XRCC1} Arg194Trp, \textit{XRCC1} Arg399Gln, \textit{OGG1} Ser326Cys and \textit{XPD} Lys751Gln, respectively (Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005a, Terry et al. 2004).

Genotyping (Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005a, Terry et al. 2004) was carried out via an AcyclPrime FP SNP detection kit, which included dideoxynucleotide triphosphates labeled with R110 or TAMRA (PerkinElmer Life Science, Boston, MA). The reaction mixture included 1 μL of G/T Terminator mix, 0.05 μL of acycloprimer enzyme, 2 μL of 10× reaction buffer, 0.5 μL (10 pmol/μL) of extension primer (for a total of 7 μL of reaction mixture), and 9.45 μL of water. Plates were read with a Perkin-Elmer Victor instrument (PerkinElmer Life Science, Boston, MA).

\textbf{Taqman Assay}

Polymorphisms in \textit{XP\textsubscript{A}} (-4A/G) and \textit{XP\textsubscript{D}} (Asp312Asn) were genotyped using the Taqman assay (Taqman PCR Core Reagent kit) (Applied Biosystems, Foster City, CA), also known as the fluorogenic 5'-nuclease assay (Crew et al. 2007) (Lee et al. 1993). The following fluorogenic oligonucleotide reverse probes were used for allele recognition: 5'-
AAGCCCCGTGGCCTGCGCCGATCTC[C/T]GGCCCACTCCGGAGGACC TAGCTCCC-3' for XPA and 5'-
CGGGGCTCACCTGCAGCAGCTT[C/T]GGCCCACTCCGAGGACCTAGCTCCC-3' for XPD. PCR amplification of 15 ng of genomic DNA was completed in a thermal cycler (95°C for 10 minutes, 40 cycles of 95°C for 25 seconds, 65°C for 1 minute) (ABI 7500, Applied Biosystems). Finally, “the fluorescence profile of each well was measured in an ABI 7500HT Sequence Detection System and the results were analyzed with Sequence Detection Software (Applied Biosystems)” (Crew et al. 2007). The Taqman assay with fluorogenic 5'-nuclease is a more robust genotyping method than FP “with well-established commercially available assays” (Crew et al. 2007) (Lee et al. 1993).

**Sequenom**

Sequenom's high-throughput matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry was used for genotyping polymorphisms in XPF (Arg415Gln) and XPG (His1104Asp) (BioServe Biotechnologies, Laurel, MD) (Ahn et al. 2004, Fannon et al. 2002). The primers used for PCR amplification of genomic DNA were 5'-
GGAAACTAGGAGGACAAGTGAG-3' and 5'-GACTTCTTCAGCTTTGCTATCC-3' for XPF, and 5'-CAGCTGTCTCCTTTGTACATTCC-3' and 5'-AAACCCAGAAGAGGGAGGCATAAC-3' for XPG. PCR was conducted using Taq polymerase (0.02 μL) (Quiagen, Valencia, CA), water (0.73 μL), 10 x Buffer B (0.5 μL) (Solis Biodyne), MgCl2 (0.5 μL, at a concentration of 25 mmol/L), dNTPs (2.5 mmol/L each, 0.25 μL), the forward and reverse primers described above, and 2 μL of DNA, which was diluted to 2.5 ng/μL (total volume = 5 μL) (Ahn et al. 2004). The Sequenom method is the most recent genotyping method used in the LIBCSP and was chosen for its ability to multiplex (Crew et al. 2007).
Quality Control

As quality control measures, "controls for genotype at each locus and two no DNA controls were included on each plate" and "any samples that were outside the variables defined by the controls were identified as non-informative and retested" (Crew et al. 2007). The positive controls for genotype were obtained through direct sequencing of relevant SNPs (Shen et al. 2005a). In addition, duplicates for approximately 8-11% of DNA samples, depending on the polymorphism, were selected for re-sampling (Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005a, Terry et al. 2004). Laboratory personnel were blinded to both case/control status and duplicate sample status. Duplicate samples for all polymorphisms showed concordance of greater than 90% based on the $\kappa$ statistic (Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005a, Terry et al. 2004). For discordant pairs, data from the first genotyping assay were used, though including or excluding these pairs did not meaningfully alter overall results (Terry et al. 2004) (Dr. Mary Beth Terry, personal communication 2012).

COVARIATE ASSESSMENT

As described previously (Chapter II), trained interviewers completed two-hour in home questionnaires with both cases and controls, which evaluated information on a wide range of potential confounders, effect modifiers and known or suspected breast cancer risk factors (Gammon et al. 2002b). Most cases and controls also completed a self-administered validated Block FFQ (Block et al. 1986, 1990, Gammon et al. 2002b, Potischman et al. 1997). Also, most cases signed a medical release form allowing access to information on clinical characteristics of their disease (Gammon et al. 2002b)

Potential confounders for this investigation include age in five-year groupings, educational attainment (less than high school, high school graduate, some college, college
graduate, or post college), religion (none, Protestant, Catholic, Jewish, other), pre-tax annual household income for the year prior to the reference date (<$15,000, $15,000-$19,999, $20,000-$24,999, $25,000-$34.99, $35,000-$49,999, $50,000-$69,999, $70,000-$89,999, or $90,000+), parity (continuous), age at first birth (continuous), and race (white, African American, or other). Information on all of these variables was collected in the main case-control questionnaire (Gammon et al. 2002b). Potential non-genetic effect modifiers are menopausal status (postmenopausal vs. premenopausal), and fruit or vegetable intake (high vs. low intake).

The fruit and vegetable consumption variable was derived from an FFQ assessing dietary patterns in the year prior to the reference date (Block et al. 1986, 1990, Gammon et al. 2002b, Potischman et al. 1997). Menopausal status was defined using self-reported information regarding date of last menstrual period, smoking status, hormone replacement therapy use, and history of hysterectomy or oophorectomy (Gammon et al. 2002b). A woman was considered postmenopausal if she reported a complete oophorectomy prior to reference date or that her last menstrual period preceded the reference date by at least 6 months. More detailed information regarding variable definition for menopausal status can be found in Chapter II.

This dissertation aims to evaluate associations between traffic PAHs and breast cancer divided into subtypes based on tumor markers (specifically, with respect to hormone receptor status and \( p53 \) mutation status). Hormone receptor (ER/PR) status was ascertained for LIBCSP cases by abstracting their pathology reports (Gammon et al. 2002b). Prior publications report that the majority of LIBCSP cases have ER+/PR+ breast tumors (e.g. Gammon et al. 2004b, Steck et al. 2007). Breast tumor \( p53 \) mutation status was analyzed using archived paraffin-embedded tumor tissue (n = 859). The extracted tumor DNA was analyzed for mutations in exons 5-8 of the \( p53 \) gene by a multistep process wherein samples were amplified using PCR and
screened via the Surveyor Mutation Detection Kit (Transgenomic, Omaha, NE, USA) (Rossner et al. 2008). Samples screening positive for possible mutations were confirmed and identified through PCR amplification and direct sequencing with an ABI 3100 capillary sequencer (Applied Biosystems Inc, Foster City, CA, USA) (Mordukhovich et al. 2009, Rossner et al. 2008). A total of 151 p53 mutations in 15% of analyzed tumors were detected (83% point mutations, 17% insertions/deletions) (Morduhovich et al. 2009, Rossner et al. 2008). Case characteristics did not significantly vary by tumor tissue availability (Rossner et al. 2008).

The LIBCSP reports associations between overall breast cancer risk and parity and age at first birth in age-adjusted models (Gammon et al. 2002b). Null results were found for education, income, and race in age-adjusted models (Gammon et al. 2002b). More detailed analyses in the LIBCSP have revealed positive associations between breast cancer and reduced fruit and vegetable intake (Gaudet et al. 2004). The LIBCSP also reported a strengthened negative association between fruit and vegetable intake and postmenopausal breast cancer among women with ER-positive tumors (Gaudet et al. 2004). No other potential confounders or effect modifiers have been evaluated with respect to tumor p53 mutation or hormone receptor status.

STATISTICAL APPROACH

Descriptive Statistics

The sample size, mean, median, range and interquartile range were calculated for traffic PAH variables from all analyses, including sensitivity analyses (see Chapter III). Means and sample sizes for all exposure variables were also calculated according to case-control status. Spearman and Pearson correlations (Rothman and Greenland 1998) between selected exposure variables were evaluated as well (Chapter III). All statistical analyses were conducted using SAS versions 9.1.3 or 9.3 (SAS Institute, Cary, NC).
Exposure Variable Considerations

Two residential traffic PAH exposure variables were chosen as the main a priori exposures of interest: 1995 exposure (complete case analysis) and 1980-1995 exposure (with up to 20% multiple imputation). These variables were selected in order to maximize study power, while simultaneously benefiting from the availability of long-term data in the LIBCSP. Associations between breast cancer and longer-term traffic PAH exposure durations were explored as well (1960-1990, 1960-1995). The association between breast cancer and peak (i.e. maximum annual) exposure between 1960 and 1990 was evaluated in order to explore the importance of high short-term traffic PAH exposure relative to cumulative long-term exposure (Dr. Jan Beyea, personal communication 2009).

Some researchers believe that environmental exposures may be most relevant to breast cancer risk during certain critical windows of exposure, such as prior to first birth or during the teenage years (Hiatt et al. 2009). The LIBCSP study likely does not have sufficient power to effectively explore this possibility. However, prior LIBCSP investigations report associations between breast cancer and long-term cumulative measures of PAH-related exposures, such as long-term residential passive smoking (Gammon et al. 2004b) and lifetime intake of grilled or smoked meat (Steck et al. 2007). A 2006 LIBCSP study explored the possibility of critical windows of susceptibility with respect to average lifetime alcohol intake and reported that associations with breast cancer did not vary by timing of exposure (Terry et al. 2006).

This analysis utilized relative rather than absolute benzo[a]pyrene exposure estimates (Dr. Jan Beyea, personal communication 2011). Specifically, the value for long-term PAH exposure variables was calculated as a function of average estimated benzo[a]pyrene exposure for the year 1995. An absolute benzo[a]pyrene exposure calculation may be misleading, as benzo[a]pyrene is
meant to be a surrogate for all traffic PAHs (Dr. Jan Beyea, personal communication 2011).

**Main Effects Regressions**

Cubic smoothing b-splines (Hastie et al. 2009, Rothman and Greenland 1998) were implemented in order to derive a data-driven exposure-outcome curve for the unimputed 1995, 1980-1995, and 1960-1990 traffic exposure variables in relation to breast cancer risk. Spline analyses determined that the use of a continuous exposure variable was not appropriate (see Chapter III), and that the use of categorical variables was more fitting. Quantile cutpoints for the exposure variables (<50th (referent), 50-<75th, 75-95th, and >95th percentiles) were informed by the spline figures (see Chapter III for more details).

Associations between residential vehicular traffic exposure estimates and breast cancer case-control status were evaluated using multivariable unconditional logistic regression to estimate odds ratios (ORs) and 95% confidence intervals (95% CIs) (Hosmer and Lemeshow 1989). For the years spanning 1980 and 1995, the partially imputed traffic PAH exposure variable included women with up to 20% of their total traffic PAH exposure originally missing during these years. Exposure gaps were filled using multiple imputation (Rubin 1996), described in detail in Chapter II. Participants with more than 20% missing data for this time span were dropped from the analysis. For the variable representing exposure during the year 1995 only, a complete case analysis was employed.

Analyses were also conducted among the following subsets of participants: (1) a complete case analysis among women with no missing exposure information between the years 1980 and 1995, (2) among women with up to 30% of their residential traffic PAH exposure information imputed between the years 1980 and 1995, (3) among a subset of women with PAH exposure data spanning the years 1960 to 1995 (up to 20% multiple imputation), and (4) among a
subset of women with PAH exposure data spanning the years 1960 to 1990, in order to account for possible latency effects (up to 20% multiple imputation). Among women with up to 20% of imputed exposure information between 1960 and 1990, the association between breast cancer and peak annual benzo[a]pyrene exposure was explored as well (Hosmer and Lemeshow 1989, Rothman and Greenland 1998).

In addition, the following sensitivity analyses were conducted. The association between breast cancer and traffic exposure in 1995 was evaluated among women residing in their current home for 15 or more years (Appendix Table A2.2), and participants with the highest exposure levels (top 2 and 5) were removed from main effects regression analyses for the following variables: 1995, 1980-1995, 1960-1990 (Appendix Table A2.3). Finally, additional sensitivity analyses were conducted for the 1960-1990 variable (single imputation): varying the relative contribution of the pre-arrival surrogate to total dose (Appendix Table A2.4), removing the contribution of intersections from the exposure model (Appendix Table A2.5), and imputing according to census place (Appendix Table A2.6).

Associations between breast cancer and 1995 and 1960-1990 traffic PAH exposure levels were also explored with breast cancer cases subdivided according to tumor characteristics. Specifically, this study evaluated associations with breast cancer subtypes categorized according to tumor $p53$ mutation status ($p53$-mutation negative vs. $p53$-mutation positive) and hormone receptor status. Tumor hormone receptor status was classified in two different ways for this subanalysis: (1) ER+/PR+ versus ER-/PR+, ER+/PR-, or ER-/PR-, and (2) ER-/PR- versus ER+/PR+, ER+/PR-, or ER-/PR+ (Potter et al. 1995, Yang et al, 2011). The first categorization was chosen for public health reasons: the ER+/PR+ subtype is the most common breast cancer subtype among US women and may account for the higher levels of breast cancer incidence in
this country relative to lower-risk countries (Yasui and Potter 1995). The second categorization was chosen because there is reason to believe that ER-/PR- breast tumors differ in etiology from luminal ER+ or PR+ cancers (Potter et al. 1995). Effect estimates for different tumor subtypes were compared using ratios of the ORs (Schlesselman 1982). The 1995 exposure variable was chosen in order to maximize power for tumor subgroup analyses, and the 1960-1990 variable was chosen because of the information it provides on long-term exposures. The same quantile cutpoints used for the main effects analyses were used for the 1995 variable when evaluating tumor heterogeneity by hormone receptor status (see Chapter 3). The two upper quantiles were collapsed due to sample size constraints within tumor subtype subgroups when evaluating this association for the 1960-1990 variable, and when examining tumor heterogeneity by \( p53 \) mutation status (Appendix Table A2.7).

**Confounding**

Because cases and controls were frequency matched by five-year age group, all statistical models must be adjusted for age group at reference (age at date of diagnosis for cases and age at date of identification for controls). Other potential confounders were identified by a thorough review of the relevant literature and analysis of causal diagrams (specifically, directed acyclic graphs) (Shrier and Platt 2008). \textit{A priori}, it was decided that a potential confounder would be included in regressions if it caused at least a 10% change in the main effect of the 1995 traffic PAH variable on breast cancer risk when added to age-adjusted models (Rothman and Greenland 1998; see Chapter III). Potential confounders include educational level (less than high school, high school graduate, some college, college graduate, post college), annual household income (<$15,000, $15,000-$19,999, $20,000-$24,999, $25,000-$34,999, $35,000-$49,999, $50,000-$69,999, $70,000-$89,999, $90,000+), body mass index at reference (in kg/m\(^2\)), race (White,
African American, other), religion (none, Protestant, Catholic, Jewish, other), number of live births (continuous), and age at first birth (continuous).

The amount of missing data was low for most of these variables (n=10 for education, n=3 for race, n=0 for parity, and n=1 for age at first birth, out of a total of 3,064 study participants; Gammon et al. 2002b). However, 11.5% of participants were missing information on household income (Gammon et al. 2004a). Therefore, income values were imputed from regression models for cases and controls. These models included age, race and education (Gammon et al. 2004a, Rubin 1996).

**Effect Modification**

**DNA Repair Variants**

This analysis examines gene-environment interactions between traffic benzo[a]pyrene exposure (years 1995 and 1960-1990) and genetic variants in NER and BER DNA repair pathways (for ERCC1, OGG1, XPA, XPD, XPF, XPG, and XRCC1) with respect to breast cancer risk. The PAH exposure variable for the year 1995 was used in order to maximize statistical power for gene-environment interaction analyses. However, the 1960-1990 variable may be a more accurate reflection of true long-term exposure opportunity, and was therefore examined as well. All of the genetic variants have been previously tested for deviations from HWE within the LIBCSP study population, and all were found to be in HWE among controls (Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005, Terry et al. 2004).

For most genetic variants of interest, gene-environment interactions were explored using a co-dominant genetic model in which participants with heterozygote and variant homozygote genotypes are assessed independently relative to a common referent group comprising women with dominant homozygote genotypes (Ziegler and Konig 2006). This approach was chosen
because there is no \textit{a priori} reason to combine women with heterozygote and variant homozygote genotypes into a single group, and because previous LIBCSP reports assessing gene-environment interactions between DNA repair polymorphisms and PAH-related exposures suggested differences between participants with heterozygote and variant homozygote genotypes that were not present when evaluating the main effects of these polymorphisms on breast cancer risk (Crew et al. 2007, Terry et al. 2004). However, a dominant genetic model was implemented when any cell sizes comprised fewer than 10 participants. Specifically, in these cases, women with heterozygote and variant homozygote genotypes were assessed relative to a referent group comprising women with dominant homozygote genotypes (Ziegler and Konig 2006). Effect modification by genetic polymorphisms was evaluated using two methods: (1) running separate regression models stratified by genotype and (2) evaluating multiplicative interactions using likelihood ratio tests (Rothman and Greenland 1998, Ziegler and Konig 2006).

In addition, effect modification by total number of 'high-risk' alleles was evaluated via (1) analyses stratified by low, intermediate, and high-risk genotype categories (Crew et al. 2007), and (2) likelihood ratio tests to assess multiplicative interactions (Rothman and Greenland 1998). 'High-risk' alleles may be variant or wild-type alleles, and were determined by examining associations between traffic PAHs and breast cancer within models stratified by genotype (see Chapter IV). Previously reported main effects of genotypes on breast cancer odds are shown in Appendix Table A2.8.

\textbf{Non-Genetic Covariates}

Interactions between traffic PAH exposure variables (1995 [CCA] and 1960-1990 [20\% MI]) and non-genetic covariates were evaluated with respect to breast cancer occurrence. Specifically, potential effect modification by menopausal status (postmenopausal vs.
premenopausal; Appendix Table A2.9) and fruit and vegetable intake (high vs. low, classified according to previously reported main effects on breast cancer risk, Gaudet et al. 2004) were examined (see Chapter III). Interactions were evaluated by running stratified regressions as well as through likelihood ratio tests (Rothman and Greenland 1998).

Menopausal status was initially classified as unknown for the 11.8% of women who had a hysterectomy but not an accompanying complete oophorectomy or who were using hormone replacement therapy. Information on smoking status was used to reduce this percentage such that "any smoker with unknown menopausal status was categorized as postmenopausal if her age was >54.8 years (90th percentile for natural menopause among smoking controls), and any nonsmoker with unknown menopausal status was categorized as postmenopausal if her age was <55.4 years (90% percentile for natural menopause among nonsmoking controls)" (Gammon et al. 2002b). This reduced the percentage of participants with missing menopausal status to 3.04% (Gammon et al. 2002b).

**Interpretation of Results**

This investigation did not rely solely on significance testing for interpreting main effects of traffic PAHs on breast cancer risk (overall or with the breast cancer outcome categorized according to tumor $p53$ mutation status or hormone receptor status), or when evaluating effect modification by genes and other covariates. The precision of confidence intervals, the sizes of effect estimates, and patterns of associations were also taken into account when interpreting results (see Chapters III and IV). Results from this investigation, whether positive or null, were carefully interpreted in light of study limitations and strengths. The role of chance findings was acknowledged when discussing results.
STRENGTHS AND LIMITATIONS

Traffic PAH Exposure Model

Data regarding tailpipe emissions and transportation patterns are not reliably available prior to the year 1960. Therefore, residential traffic PAH exposures were not modeled prior to this date (Beyea et al. 2005). In addition, the LIBCSP did not collect residential history information outside of Nassau and Suffolk counties or for addresses within Nassau or Suffolk counties at which a participant resided for less than one year. Non-residential traffic-related PAH exposure estimates, such as while at work or driving, could not be modeled. Exposures while driving could be especially important for women living in the lowest exposure regions in Nassau or Suffolk counties (Beyea et al. 2005). Calculated "peak exposures on roads were generally ten times higher than average residential exposures" (Beyea et al. 2005). These omissions inevitably lead to some degree of misclassification. The effect on model accuracy of time spent away from home may, however, be less than expected due to the large exposure component at night arising from pollutant drift from rush-hour traffic (Beyea et al. 2005). Other limitations of the exposure model include uncertainties in chosen model parameters, such as dispersion variables, meteorological variables, the relative importance of emissions at traffic intersections, or building penetration factors, which could result in misclassification of the traffic PAH exposure variable.

The exposure model does not incorporate non-traffic sources of PAHs, including ambient sources such as cigarette smoke, residential heating, cooking, or leaf-burning. However, traffic is often the largest ambient source of PAHs in or near urban areas (Bostrom et al. 2002, Dubowsky et al. 1999, Dunbar et al. 2001, IARC 1989), and is usually the largest source of air pollution in these areas (Maitre et al. 2002, WHO 1996). Furthermore, traffic is more likely to release smaller, more health-relevant particles than some of the aforementioned ambient
exposure sources (Kocbach et al. 2006). The geographic model also did not incorporate information regarding 'canyon effects' due to row housing (Raaschou-Nielsen et al. 2000), which is rare in the study area, and did not account for the small number of roads that were added to the study area after 1960, which is likely to have a minor impact on exposure estimates based on sensitivity analyses (Beyea et al. 2005).

Only benzo[a]pyrene was modeled as a surrogate for all traffic PAHs. Traffic pollution also contains a large number of non-PAH chemicals, some of which are confirmed carcinogens (e.g. benzene and formaldehyde; Bostrom et al. 2002, IARC 1989, 2006, 2008). Therefore, it is impossible to definitively ascribe health effects to a single traffic pollutant based on an observational study. However, benzo[a]pyrene is one of the most carcinogenic PAHs and is representative of overall PAH exposure from vehicle exhaust (Bostrom et al. 2002, Fertmann et al. 2002). Furthermore, due to their "potency and high concentrations" in vehicular exhaust, particulate-phase PAHs are considered among the most carcinogenic components of traffic emissions (Boothe and Shendell 2008). Finally, despite being correlated, traffic-related pollutants do not have identical distributions. In the LIBCSP, exposure estimates were modeled using PAH-specific vehicle emissions data, PAH-specific dispersion parameters, and were validated and calibrated against residential soil PAH levels and PAH-DNA adduct levels in mononuclear cells (Beyea et al. 2006). As an illustrative example, the benzo[a]pyrene geographic model used in the LIBCSP needed to be altered in order to predict carbon monoxide levels for a validation exercise. Specifically, the carbon monoxide model excluded "all depletion phenomena, because deposition, washout, and photo decay are negligible in the case of CO" (Beyea et al. 2006).

The above limitations should be interpreted in light of realistic standards in the literature.
Specifically, the standard approach in environmental epidemiology is to assign exposure estimates (often short-term and/or self-reported) based on relatively crude measures such as readings from sparse pollutant monitors or distance from major roads, and thus even uncertain modeling parameters are likely to be an improvement over these approaches (Beyea 1999). The state of the art exposure model used to generate LIBCSP residential traffic PAH exposure estimates is a vast improvement over many comparable studies. Model parameters were validated and calibrated against extensive PAH-related field data from the LIBCSP, thereby greatly increasing confidence in their credibility (Beyea et al. 2006). Sensitivity analyses suggest that modeling uncertainties are not likely to strongly impact residential traffic PAH exposure estimates (Beyea et al. 2005). Most LIBCSP participants provided at least one address that geocoded to street level (for addresses through the year 1990, n= 2,601; for 1995 addresses, n = 2,655).

Specific strengths of the traffic-related PAH exposure model include a comprehensive individualized ambient benzo[a]pyrene exposure metric spanning a large geographic area, and accounting for participants' residential histories for up to a 35 year period. Exposure estimates were derived using a sophisticated historical geographic model that incorporated information on a wide range of factors including changing historical vehicular emissions and transportation patterns, meteorologic data, pollutant dispersion and deposition variables, acceleration and deceleration at intersections, cruise and cold-engine emissions, building penetration factors, and background ambient PAH from distant traffic and other PAH sources (Beyea et al. 2006). The model was validated and calibrated against field data (specifically, residential soil PAHs, PAH-DNA adducts in mononuclear cells, and carbon monoxide concentrations at an EPA monitoring site; Beyea et al. 2006). Participants were exposed to a wide range of residential traffic PAH
levels, due to the gradient of urbanization across the study area (Beyea et al. 2006). Finally, since breast cancer develops over many years (Clark et al. 1997), a long modeled exposure period may be relevant to assessing relationships with breast cancer. Residential PAH and other pollutant exposure levels can vary greatly from year to year for a given woman due to, for example, changes in emissions or moving to a new residence (Beyea et al. 2005).

**Other Considerations**

This investigation may be subject to potential limitations standard to epidemiological studies such as laboratory errors with respect to genotyping or tumor classification, inaccurate reporting or recording of confounding or modifying covariates, and differential misclassification with respect to case-control status (Rothman and Greenland 1998). However, the LIBCSP employed extensive quality control measures, both with regard to genotyping (e.g. genotyping randomly distributed duplicate DNA samples, Crew et al. 2007; see Chapter II for more details) and other data collection and storage procedures (e.g. running data range and logic checks and re-interviewing a random sample of 20% of participants, Gammon et al. 2002b; see Chapter II for more details). Identical data collection and storage strategies were used for case and control participants. These measures likely reduce the probability and extent of potential study limitations (Gammon et al. 2002b).

Study participation rates differed between eligible cases and controls (82.1% vs. 62.7%, respectively) (Gammon et al. 2002b). Reduced participation among eligible controls was driven by lower participation rates among elderly women (Gammon et al. 2002b). Finally, the LIBCSP is a fairly homogenous study population with respect to race (primarily white), socioeconomic status (relatively high) and geographic area. This is both a strength and a limitation. Participant homogeneity is an advantage with respect to confounder control as it decreases the likelihood of
residual confounding, but participant homogeneity also reduces the generalizability of study findings.

The LIBCSP provides a large study population with high response rates for both cases and controls (Gammon et al. 2002b), and DNA was collected from most LIBCSP participants (Gammon et al. 2002b). The higher resulting study power is important for effectively evaluating associations between traffic PAHs and breast cancer, and is especially important for evaluating statistical interactions, such as with genetic variants. A case-control study provides the most efficient means of examining associations with breast cancer risk, as a prohibitively large and resource-intensive cohort would be needed to obtain the same number of cases due to the relatively low population incidence of breast cancer (ACS 2009-2010, Rothman and Greenland 1998). The LIBCSP utilized extensive quality control procedures for questionnaire data and for the laboratory component of the investigation, as described above (Crew et al. 2007, Gammon et al. 2002b). The LIBCSP also recruited incident rather than prevalent cases, which increases the likelihood of identifying risk factors for breast cancer occurrence rather than survival (Gammon et al. 2004b, Rothman and Greenland 1998).

Additional strengths of this study include access to the resources of a large population-based investigation with extensive information on a wide range of variables including demographic and lifestyle factors, active and passive cigarette smoking history, reproductive and medical history, dietary intake, tumor markers and subtypes, and a variety of genetic polymorphisms (Gammon et al. 2002b). This facilitates effectively controlling for potential confounders, assessing effect modification, and evaluating tumor subtype-specific risk factor associations.
SUMMARY

The Long Island Breast Cancer Study Project is a large, population-based case-control study which has developed a unique and comprehensive long-term measure of residential exposure to ambient PAHs from vehicular traffic (Beyea et al. 2006). The historical exposure model was validated against relevant field data from the LIBCSP (specifically, soil PAHs, PAH-DNA adduct levels in mononuclear cells, and carbon monoxide concentrations at an EPA monitoring location) (Beyea et al. 2006). The LIBCSP has collected extensive information regarding breast tumor subtypes and a wide range of established or possible breast cancer risk factors, confounders, and effect modifiers, including genetic polymorphisms in relevant biological pathways (Gammon et al. 2002b, Gammon and Santella 2008).

This dissertation analysis aims to (1) evaluate associations between residential traffic-related PAH exposure and breast cancer, and (2) assess interactions between traffic PAH exposure and DNA repair polymorphisms in nucleotide excision repair and base excision repair pathways with respect to breast cancer risk. Secondary research goals include evaluating associations between traffic PAHs and breast cancer categorized according to tumor subtypes (based on hormone receptor status and \(p53\) mutation status), and assessing interactions between traffic PAHs and fruit and vegetable intake as well as menopausal status with respect to breast cancer occurrence.

PAHs are ubiquitous environmental carcinogens which are found in high concentrations in traffic exhaust (Boothe and Shendell 2008, IARC 2010). PAHs are potent mammary carcinogens in laboratory animals (el-Bayoumy et al. 1995), and PAH-DNA adducts, air pollution, and traffic exposure are consistently linked with breast cancer in epidemiological studies (Beyea et al. 2006, Bonner et al. 2005, Gammon and Santella 2008, Rundle et al. 2002a).
However, the relevant epidemiological literature to date is sparse and no previous research has attempted to evaluate the association between such a comprehensive and long-term air pollution or traffic exposure metric and breast cancer risk, or to evaluate interactions between air pollution, DNA repair genotypes, and fruit and vegetable intake.
CHAPTER III: EXPOSURE TO TRAFFIC-RELATED POLYCYCLIC AROMATIC HYDROCARBONS AND BREAST CANCER RISK

INTRODUCTION

Breast cancer is the most common malignancy among women in the United States (ACS 2011-2012). Polycyclic aromatic hydrocarbons (PAHs) are ubiquitous pollutants formed from incomplete combustion (Boström et al. 2002). PAHs are genotoxic prooxidants, confirmed human lung carcinogens, and potent mammary carcinogens in laboratory animals (IARC 2010). However, the association between PAHs and breast cancer in women is unclear (Gammon and Santella 2008). Previous population studies have reported associations between PAH-related exposures and breast cancer. For example, PAHs bind to DNA, including in breast tissue, and the resulting DNA adducts have been linked with breast cancer in epidemiological studies, although the research is scant (Gammon et al. 2002c, Rundle et al. 2000). PAH-DNA adducts reflect short-term exposures (Gammon et al. 2002c), whereas breast cancer is thought to develop over many years. Thus, it is of interest to evaluate longer-term PAH exposures in relation to breast cancer risk.

Vehicular traffic is a major source of PAH exposure, especially near urban areas (Fromme et al. 2004), and is a ubiquitous and generalizable air pollution exposure surrogate (Boström et al. 2002). Studies consistently report positive associations between breast cancer and air pollution, although the exposure assessment methods in these previous reports varied (Bonner et al. 2005, Crouse et al. 2010, Lewis-Michel et al. 1996, Nie et al. 2007, Raaschou-Nielsen et al. 2011). Some investigations relied on simple traffic density data or sparse monitors, evaluated relatively brief periods of exposure, focused on nitrogen oxides rather than carcinogenic
particulate pollution, or used unvalidated exposure assessment methods. Further progress requires sophisticated modeling to reconstruct long-term cumulative exposures to ambient PAHs.

Breast cancer risk factor profiles differ by menopausal status and tumor subtype (ACS 2011-2012, Chen and Colditz 2007), and fruits and vegetables may modify the carcinogenic effects of PAHs via antioxidant and other chemopreventive properties (Hecht et al. 2000, Jin et al. 2006). However, whether the association between ambient PAHs and breast cancer varies by fruit/vegetable intake, tumor characteristics, and menopausal status is not well understood.

Our population-based study aims to evaluate the association between breast cancer incidence and traffic-related PAH exposure, overall and within subgroups of fruit/vegetable intake, menopausal status and tumor subtypes (Gammon et al. 2002b). For our study reported here, we utilized long-term, individualized residential traffic PAH exposure estimates, which were reconstructed using a comprehensive, extensively validated exposure model (Beyea et al. 2006). It is important to clarify the association between traffic PAHs and breast cancer given the high incidence of breast cancer and widespread exposure to traffic pollution worldwide.

MATERIALS AND METHODS

We used data from the case-control component of the Long Island Breast Cancer Study Project (LIBCSP), a population-based investigation conducted among women residing in Nassau and Suffolk counties in Long Island, NY (Gammon et al. 2002b). All participating institutions provided Institutional Review Board approval for this study.

Study Population

Eligible case participants were diagnosed with a first primary invasive or in situ breast cancer between August 1996 and July 1997, and identified via rapid case ascertainment through contact with local pathology departments (Gammon et al. 2002b). Eligible control participants
were women with no history of breast cancer, and were identified using random digit dialing (Waksberg 1978) for women under age 65 and Health Care Finance Administration records for women aged 65 years or older. Controls were frequency matched based on the expected five-year age distribution among the case participants.

Respondents included 1,508 cases and 1,556 controls (82.1% and 62.7% of eligible participants, respectively), who ranged between 20 and 98 years of age and were mostly postmenopausal (67.4%) and white (92.8%); the racial distribution reflects that of the study counties at the time of data collection (Gammon et al. 2002b). Over 50% of participants reported a household income of $50,000 or greater in the year prior to the study interview (Gammon et al. 2002b). In previous LIBCSP reports, we found that increased breast cancer risk was associated with: early age at menarche, few or no births, and little or no breastfeeding (Gammon et al. 2002b, Shantakumar 2007); increased body size (Eng et al. 2005) and little or no physical activity (McCullough et al. 2012) among postmenopausal women; low fruit/vegetable intake (Gaudet et al. 2004) and low flavonoid intake (Fink et al. 2007); and increased blood levels of PAH-DNA adducts (Gammon et al. 2002c), long-term residential environmental tobacco smoke exposure (Gammon et al. 2004b), and increased grilled/smoked food intake (Steck et al. 2007).

**Study Questionnaire**

Trained interviewers administered an at-home, two-hour structured questionnaire within a few months of diagnosis (cases) or study recruitment (controls). Detailed questionnaire information includes residential history, household income, educational attainment, race, religion, reproductive and medical histories, smoking, exogenous hormone use, and body size (Gammon et al. 2002b). A validated Block food frequency questionnaire (Block et al. 1986, 1990) was self-completed by 98% of cases and controls, and assessed usual dietary intake during
the year prior to the interview, including intake of fruits, fruit juices, and vegetables (Gaudet et al. 2004).

**Tumor Subtypes**

Medical records were abstracted for 1,402 case participants to ascertain tumor estrogen and progesterone receptor (ER/PR) status subtypes (Gammon et al. 2002b). Tumor \( p53 \) mutations, which can be induced by PAHs (Greenblatt et al. 1994), were determined using archived, paraffin-embedded tissue (n=859) (Rossner et al. 2009). Briefly, the extracted tumor DNA was amplified using PCR, screened via the Surveyor Mutation Detection Kit (Transgenomic, Omaha, NE, USA), and possible mutations were confirmed with an ABI 3100 capillary sequencer (Applied Biosystems Inc, Foster City, CA, USA).

**Traffic Exposure Assessment**

Interviewers collected information regarding participants' lifetime residential histories in Nassau and Suffolk counties. Addresses at which a woman resided for at least one year were recorded and geocoded using BLR software (BLR Data, Inc. Lebanon, NH). Most participants (n=2,655) provided at least one address that geocoded to the street level. The percentage of addresses that geocoded successfully ranged between 60% in the 1960s and 85% in 1995 (Beyea et al. 2005).

The traffic exposure model (Beyea et al. 2005, 2006, 2013) used a road network comprising approximately 500,000 street segments in the New York metropolitan area, which includes Nassau and Suffolk counties. Emissions for each street segment were calculated as the product of hourly roadway-specific traffic counts and average U.S. tailpipe emissions for the years 1960, 1970, 1980, and 1990. Estimates for other years between 1960 and 1995 were derived via interpolation or extrapolation. Vehicular traffic estimates include adjustment for
acceleration/deceleration at intersections (Beyea et al. 2006). Emissions within 80 kilometers of
the study area were directly calculated, and the exposure model also included a background term
(proportional to exposures from the 22 counties just outside of the two-county study area) to
account for more distant roads (Beyea et al. 2006). Tailpipe emission parameters were derived
from road tunnel measurements throughout the U.S. and checked using vehicles run on
dynamometer test beds (Beyea et al. 2008). Roadway traffic counts were obtained from over
13,000 “annual average daily traffic” measurements, recorded in state and county records.

Roadway emissions were translated into predicted residential ambient benzo[a]pyrene
concentrations, considered a surrogate for exposure to all traffic PAHs, using standard
meteorological dispersion and deposition models (Beyea et al. 2006, Chock 1978, Viegele and
Head 1978). Up to 100 meters from a road, traffic PAH levels were estimated using highway
line-source models (Chock 1978). Beyond 100 meters, a standard Gaussian plume model with
Briggs dispersion parameters (Viegele and Head 1978), which incorporated data on wind speed
and direction, rain washout, photo decay, and particle deposition, was used (Beyea et al. 2006).
Meteorological data were collected at Brookhaven National Laboratory (year 1993). Changing
the year (1990) or location (MacArthur Airport) of meteorological data collection for a
sensitivity analysis did not appreciably alter exposure estimates (Beyea et al. 2005). Finally,
indoor exposures were calculated using a building penetration factor of 0.75 (Long et al. 2001).

Model Validation

The exposure model was modified to predict measured residential soil and carpet PAH
levels among women living at their current address for at least 15 years, PAH-DNA adduct levels
in circulating mononuclear cells collected at interview, and monitoring data for carbon
monoxide, a traffic-related pollutant (Beyea et al. 2006). The model accurately predicted soil
PAHs, PAH-DNA adduct levels, and carbon monoxide concentrations, but not PAH levels in carpet dust (Beyea et al. 2006). The latter finding is consistent with a study showing that residential ambient PAH concentrations do not correlate with house dust PAH levels (Fromme et al. 2004).

The model was calibrated against soil and PAH-DNA adduct data through chi-squared minimization (Beyea et al. 2006). The soil-calibrated model provided the best fit to field data and was therefore used for regressions against breast cancer risk. Emissions during engine warm-up, which were included during model building, did not contribute to predicting soil PAH levels in validation exercises and were dropped from the exposure model (Beyea et al. 2006).

**Missing Exposure Data**

Residential histories during 1960-1995 were incomplete for study participants who migrated to Long Island from other locations or whose recollection of prior addresses was incomplete (Beyea et al. 2013). Thus, we used a surrogate for pre-arrival exposure, which was standardized to the average estimated traffic PAH concentrations on Long Island during the relevant period, based on five-year intervals; variance was added during multiple imputation (MI) (Beyea et al. 2013, Rubin 1996). For missing post-arrival addresses, multiple imputation was used to estimate the function converting roadway emissions to exposures at individual residences, using interpolation or extrapolation from a woman's known addresses (Beyea et al. 2013).

**Statistical Methods**

We evaluated cumulative vehicular traffic exposures for the years 1995 (the year prior to case ascertainment), 1980-1995, 1960-1995 and 1960-1990, along with maximum annual exposure during 1960-1990. We restricted participants' percentage of imputed exposure to 20%
of total exposure during 1980-1995, 1960-1995, and 1960-1990 (Beyea et al. 2013). For 1995 exposures, we conducted a complete case analysis (CCA), in which all participants had complete, unimputed exposure information. As a sensitivity analysis, we evaluated 1980-1995 traffic exposure as a CCA and with up to 30% MI. We selected a range of exposure durations in order to maximize study power, which was higher for recent exposures, while also benefiting from long-term cumulative exposure information.

We calculated descriptive statistics for the exposure variables, overall and by case-control status, and evaluated Spearman and Pearson correlations (Rothman and Greenland 1998) between selected exposure variables. Exposure levels were normalized to the average benzo[a]pyrene exposure in 1995, and expressed as relative units (Beyea et al. 2013). Statistical analyses were conducted using SAS version 9.3 (SAS Institute, Cary, NC).

Prior to undertaking regression analyses, we used age-adjusted cubic smoothing splines with knots at each unique exposure value (Hastie and Tibshirani 1990) to evaluate the association between traffic exposure (1995, 1980-1995, 1960-1990) and breast cancer risk. The spline figures indicated that the association was not linear and were used to generate traffic PAH quantile cutpoints for regression models: <50th (referent), 50-<75th, 75-95th, and ≥95th percentiles.

We used unconditional logistic regression (Hosmer and Lemeshow 1989) to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for associations between traffic estimates and breast cancer risk. When using partially imputed exposure variables, effect estimates and confidence limits were combined across 30 imputations using Rubin's rules (Rubin 1996).

All regression models were adjusted for the case-control frequency matching factor, five-year age group (Gammon et al. 2002b). Other potential confounders (educational level, annual household income, race, religion, parity, age at first birth, body mass index (kg/m²)) were
identified by a thorough literature review and analysis of directed acyclic graphs (Shrier and Platt 2008), and evaluated for model inclusion using a 10% change in estimate criterion relative to age-adjusted models (Rothman and Greenland 1998). No covariates meaningfully altered age-adjusted effect estimates (years 1995, 1960-1990), and thus all models adjust only for age group.

We explored effect modification by menopausal status (pre vs. postmenopausal) and fruit or vegetable intake in the year prior to the study interview (≤34 vs. 35+ servings/week) for 1995 and 1960-1990 exposures using stratified analyses and likelihood ratio tests for multiplicative interaction (Rothman and Greenland 1998). Menopausal status was defined using data from the study questionnaire (Shantakumar 2007). Fruit/vegetable intake was dichotomized based on previously reported effects on breast cancer risk in the LIBCSP (Gaudet et al. 2004).

We evaluated associations between traffic PAHs and breast tumor subtypes categorized according to p53 mutation status (mutation-positive vs. mutation-negative; Mordukhovich 2010) and hormone receptor status (ER+PR+ vs. all other subtypes, and ER-PR- vs. all other subtypes; Yang et al. 2011, Yasui and Potter 1995). Using polytomous logistic regression (Hosmer and Lemeshow 1989), we simultaneously compared cases with each individual subtype to the control participants. The ratio of the odds ratios (Schlesselman 1982) was used to evaluate heterogeneity of the effect estimates across tumor types. The upper exposure quantiles (75-95th and ≥95th percentiles) were collapsed when cell sizes comprised fewer than 10 participants.

RESULTS

The number of LIBCSP respondents for whom traffic estimates are available (520-1,274 cases; 566-1,334 controls) varied according to the exposure definition and the imputed data set (some women exceeded the limit on imputation percentage in certain imputation draws), and participants showed a wide range of exposure levels (Table 1). For example, exposures during
1995 ranged between 0.02 and 30.63 relative units, with an interquartile range of 0.64 to 1.24 and, by definition, a mean of 1.00. Mean exposure levels were consistently higher among cases (Table 2). In the year 1995, mean exposures were 1.03 and 0.97 units for cases and controls, respectively. The corresponding mean levels for cumulative exposure in 1960-1990 were 227.42 and 196.71 units of average annual exposure in the year 1995. Short-term and longer-term exposure estimates were highly correlated (Supplemental Material, Table S1).

Adjusted cubic spline figures suggested an increased breast cancer risk among women with the top 1% of traffic exposure levels (data not shown). These results are based on a small number of women. Hence, we used broader quantiles in regression models in order to stabilize the results. ORs from logistic regression models comparing participants with the top 5% of exposure levels to those with exposures below the median ranged between 0.98 (95% CI: 0.62, 1.53) for 1980-1995 exposure when we used a complete case analysis approach, and 1.44 (95% CI: 0.78, 2.68) for 1960-1990 exposure with up to 20% MI (Table 3). For recent exposure in 1995, the corresponding OR for the top 5% of exposure levels was 1.14 (95% CI: 0.80, 1.64) (Table 3). Effect estimates were strongest when considering the top 5% of long-term exposure estimates (1960-1990, 1960-1995), rather than more recent exposures, though confidence intervals were wide and included the null value (Table 3). Varying the percent of imputation allowed did not appreciably alter the results shown in Table 3. Additional sensitivity analyses (see Supplemental Material) also indicated that our findings are robust.

The traffic PAH-breast cancer association was stronger among women with low fruit/vegetable intake (Table 4). For example, among women with low fruit/vegetable intake, we observed elevated risks for the top 5% of 1995 traffic exposures compared to levels below the median (OR=1.46, 95% CI: 0.89, 2.40, p for trend=0.04). Among women with high
fruit/vegetable intake, the corresponding OR was 0.92 (95% CI: 0.53, 1.60) (p-interaction=0.01).

Traffic exposure and breast cancer risk were also more strongly associated among premenopausal women (Table 4), though there was no clear dose-response relationship. For example, the OR for the association between breast cancer and 1995 traffic exposure (75-95th percentile vs. below the median) was 1.64 (95% CI: 1.13, 2.38) among premenopausal women, but 0.80 (95% CI: 0.62, 1.02) among postmenopausal women (p-interaction=0.02).

Traffic exposure in 1995 (top 5% vs. below the median) was more strongly associated with ER-/PR-negative breast tumors (OR=1.67; 95% CI: 0.91, 3.05), rather than ER- or PR-positive tumors (ratio of the ORs=2.09, 95% CI: 1.08, 4.06) (Table 5). We observed no heterogeneity of the effect estimates for tumor p53-mutation status subtypes (data not shown).

DISCUSSION

We observed a modest positive association between high-level residential exposure to vehicular traffic PAHs and breast cancer risk in this population-based study. The effect estimate for the association was increased by 14% for recent 1995 traffic exposure and by 44% for long-term (1960-1990) exposure, although the confidence intervals were wide. The effect estimate for the association between recent traffic exposure and breast cancer was increased by 46% among women with low fruit/vegetable intake, by 67% among those with hormone receptor-negative breast tumors, and by 64% among premenopausal women. Although U.S. traffic emissions are greatly reduced from the high levels of the 1960s and 1970s (Beyea et al. 2008), our results may have public health significance given widespread exposure to traffic pollution worldwide.

We observed stronger associations between traffic exposure and breast cancer risk among women with low fruit/vegetable intake, consistent with a previous LIBCSP report which found that associations with grilled/smoked meat intake – another major PAH source – were more
pronounced among participants who consumed low levels of fruits/vegetables (Steck et al. 2007). Similarly, fruit/vegetable intake was negatively associated with PAH-DNA adduct levels in the LIBCSP (Shantakumar et al. 2005) and in an epidemiologic study conducted in Europe (Palli et al. 2000). Components of fruits and vegetables have antioxidant and other chemopreventive properties and decrease carcinogenic effects of PAHs in animals (Hecht et al. 2000, Jin et al. 2006). Ours is the first population-based breast cancer study to evaluate interactions between air pollution and nutrition, and our results require confirmation.

Traffic-breast cancer associations were stronger among premenopausal women in our study, consistent with previous LIBCSP reports regarding combined active/passive smoking exposure (Gammon et al. 2004b) and PAH-DNA adducts (Gammon et al. 2002c). Similarly, a recent systematic review concluded that associations between environmental tobacco smoke and premenopausal, but not postmenopausal, breast cancer were “consistent with causality” (Johnson et al. 2011). We also found stronger associations with traffic among women with ER-/PR-negative breast tumors, which are overrepresented among premenopausal women (Chen and Colditz 2007). Previous studies conducted in an area with high industrial emissions reported positive associations between air pollution and breast cancer that varied inconsistently by menopausal status, depending on the exposure definition used, and did not vary by tumor hormone-receptor subtype (Bonner et al. 2005, Nie et al. 2007).

Traffic exposure was not related to tumor $p53$-mutation status in our study, consistent with an LIBCSP report evaluating active and passive smoking, grilled/smoked meat intake, and PAH-DNA adducts (Mordukhovich et al. 2010). Active, but not passive, smoking was associated with $p53$-mutation positive breast cancer in a smaller population study (Conway et al. 2002),

Several mechanisms potentially underlie the observed association between traffic-
related PAHs and breast cancer risk. PAHs are genotoxic, damaging DNA primarily through formation of PAH-DNA adducts; they may also induce inflammation and oxidative stress, resulting in oxidative DNA lesions (IARC 2010). Uncorrected DNA damage can cause somatic mutations in tumor suppressor genes or proto-oncogenes, which could in turn contribute to carcinogenesis (Gammon and Santella 2008, IARC 2010).

Limitations of our study include some missing address information, which was corrected in part through multiple imputation. Our results were robust to varying the percentage of imputation allowed (0-30%). In addition, the exposure model did not generate non-residential traffic exposure estimates. A nation-wide study reported that, on average, American men and women spend almost 70% of their day inside (64.97%) or just outside (2.50%) of their residences, both overall and in a large city (Chicago, IL) (Leech et al. 2002). A Canadian study found similar results among women only (Nethery et al. 2009).

The purpose of our exposure model was to estimate vehicular traffic exposures, thus by design we did not incorporate information on non-traffic sources of ambient PAHs, such as environmental tobacco smoke, cooking or heating (Beyea et al. 2013). Traffic is often the largest source of ambient PAHs near urban areas (Fromme et al. 2004, Nielsen 1996) and generates smaller, more inflammatory particles than some other ambient sources (Kocbach et al. 2006).

Benzo[a]pyrene was modeled as a surrogate for all traffic PAHs. Traffic pollution contains many chemicals, including many PAHs and other confirmed carcinogens (Boström et al. 2002). It is impossible to ascribe health effects to a single traffic pollutant based on an observational study. However, benzo[a]pyrene is one of the most carcinogenic PAHs and is representative of overall PAH exposure from vehicle exhaust (Boström et al. 2002), and particulate PAHs are among the most carcinogenic components of traffic exhaust (Boothe and
Furthermore, traffic pollutants do not have identical distributions (Beyea et al. 2006). In our study, exposures were modeled using PAH-specific emissions data and dispersion parameters, and were validated and calibrated against soil PAH levels (Beyea et al. 2006).

Strengths of our study include long-term, individualized traffic PAH exposure estimates generated from an extensively validated exposure model that incorporated information regarding historical tailpipe emissions, local traffic patterns, meteorological conditions, pollutant dispersion, deposition and decay data, building penetration, proximity to intersections, and background PAHs (Beyea et al. 2006). Participants had a wide range of exposure levels due to the gradient of urbanization across the study area (Beyea et al. 2006). In addition, the LIBCSP is a large, population-based study with detailed information regarding many covariates, which facilitates controlling for confounders, assessing effect modification, and evaluating heterogeneity within breast cancer subtypes (Gammon et al. 2002b).

CONCLUSIONS

We observed positive associations between breast cancer risk and comparatively high levels of residential traffic PAHs, particularly among women with low fruit/vegetable intake, hormone receptor negative tumors, and premenopausal breast cancer, although the confidence intervals were wide. Future studies should examine associations between traffic exposure and breast cancer within larger or more highly exposed study populations, and should evaluate interactions with genetic variants in relevant pathways. If confirmed, our results will contribute to increased understanding of breast cancer etiology, strengthen the biological plausibility of the relationship between vehicular traffic and breast cancer, and help identify vulnerable populations.
Table 3.1. Distributions of cumulative residential vehicular traffic PAH exposure estimates.

<table>
<thead>
<tr>
<th>Traffic Exposure Years(^a)</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995 (CCA)</td>
<td>2608</td>
<td>1.00</td>
<td>0.84</td>
<td>0.02</td>
<td>30.63</td>
<td>0.64, 1.24</td>
</tr>
<tr>
<td>1980-1995 (CCA)</td>
<td>1719</td>
<td>22.22</td>
<td>18.51</td>
<td>1.78</td>
<td>389.40</td>
<td>13.96, 27.85</td>
</tr>
<tr>
<td>1980-1995 (20% MI)(^b)</td>
<td>1966-1983</td>
<td>22.21</td>
<td>18.41</td>
<td>1.78</td>
<td>501.81</td>
<td>13.80, 27.59</td>
</tr>
<tr>
<td>1980-1995 (30% MI)(^b)</td>
<td>2047-2069</td>
<td>22.06</td>
<td>18.21</td>
<td>1.78</td>
<td>501.81</td>
<td>13.74, 27.28</td>
</tr>
<tr>
<td>1960-1995 (20% MI)(^b)</td>
<td>1112-1152</td>
<td>216.64</td>
<td>178.92</td>
<td>15.84</td>
<td>8876.15</td>
<td>128.98, 256.72</td>
</tr>
<tr>
<td>1960-1990 (20% MI)(^b)</td>
<td>1095-1147</td>
<td>211.47</td>
<td>175.20</td>
<td>15.66</td>
<td>5469.61</td>
<td>127.77, 251.03</td>
</tr>
<tr>
<td>1960-1990 (20% MI, peak)(^b,c)</td>
<td>1095-1147</td>
<td>13.38</td>
<td>10.77</td>
<td>0.88</td>
<td>333.70</td>
<td>7.52, 16.34</td>
</tr>
</tbody>
</table>

CCA: complete case analysis; IQR: interquartile range; MI: multiple imputation; PAH: polycyclic aromatic hydrocarbon

\(^a\) Exposures are normalized to the mean exposure level for 1995.
\(^b\) Combined over m=30 imputations. Sample size varies across imputed data sets.
\(^c\) Maximum annual exposure.
Table 3.2. Residential traffic polycyclic aromatic hydrocarbon exposure estimates by case-control status.

<table>
<thead>
<tr>
<th>Traffic Exposure Years(^a)</th>
<th>Cases ((n = 1,508))</th>
<th>Controls ((n = 1,556))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean (IQR)</td>
</tr>
<tr>
<td>1995 (CCA)</td>
<td>1274</td>
<td>1.03 (0.62)</td>
</tr>
<tr>
<td>1980-1995 (30% MI)(^b)</td>
<td>995-1009</td>
<td>22.83 (14.10)</td>
</tr>
<tr>
<td>1960-1995 (20% MI)(^b)</td>
<td>528-561</td>
<td>232.65 (125.03)</td>
</tr>
<tr>
<td>1960-1990 (20% MI)(^b)</td>
<td>520-551</td>
<td>227.42 (125.31)</td>
</tr>
<tr>
<td>1960-1990 (20% MI, peak)(^b,c)</td>
<td>520-551</td>
<td>14.45 (9.21)</td>
</tr>
</tbody>
</table>

CCA: complete case analysis; IQR: interquartile range; MI: multiple imputation
\(a\) Exposures are normalized to the mean exposure level for 1995.
\(b\) Combined over \(m=30\) imputations. Sample size varies across imputed data sets.
\(c\) Maximum annual exposure.
Table 3.3. Associations between varying time ranges of exposure to PAHs from residential traffic and breast cancer risk.

<table>
<thead>
<tr>
<th>Traffic Exposure Years</th>
<th>Cases</th>
<th>Controls</th>
<th>Percentiles of Exposure</th>
<th>Age-adjusted OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995 (CCA)</td>
<td>645</td>
<td>659</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td></td>
<td>299</td>
<td>353</td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.87 (0.72, 1.04)</td>
</tr>
<tr>
<td></td>
<td>260</td>
<td>261</td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.01 (0.82, 1.23)</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>61</td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.14 (0.80, 1.64)</td>
</tr>
<tr>
<td>1980-1995 (CCA)</td>
<td>431</td>
<td>428</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td></td>
<td>201</td>
<td>230</td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.86 (0.68, 1.08)</td>
</tr>
<tr>
<td></td>
<td>171</td>
<td>173</td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.95 (0.74, 1.22)</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>42</td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.98 (0.62, 1.53)</td>
</tr>
<tr>
<td>1980-1995 (20% MI)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>491-502</td>
<td>498-510</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td></td>
<td>225-231</td>
<td>256-264</td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.87 (0.70, 1.09)</td>
</tr>
<tr>
<td></td>
<td>192-195</td>
<td>202-205</td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.93 (0.74, 1.18)</td>
</tr>
<tr>
<td></td>
<td>48-49</td>
<td>43-45</td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.04 (0.68, 1.61)</td>
</tr>
<tr>
<td>1980-1995 (30% MI)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>514-526</td>
<td>526-539</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td></td>
<td>230-238</td>
<td>266-274</td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.87 (0.70, 1.08)</td>
</tr>
<tr>
<td></td>
<td>195-201</td>
<td>205-209</td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.95 (0.75, 1.20)</td>
</tr>
<tr>
<td></td>
<td>49-50</td>
<td>45-46</td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.05 (0.69, 1.61)</td>
</tr>
<tr>
<td>1960-1995 (20% MI)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>267-301</td>
<td>299-327</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td></td>
<td>123-137</td>
<td>134-158</td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.99 (0.72, 1.35)</td>
</tr>
<tr>
<td></td>
<td>97-110</td>
<td>110-120</td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.92 (0.66, 1.29)</td>
</tr>
<tr>
<td></td>
<td>24-29</td>
<td>19-22</td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.38 (0.74, 2.55)</td>
</tr>
<tr>
<td>1960-1990 (20% MI)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>262-287</td>
<td>289-320</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td></td>
<td>122-139</td>
<td>136-155</td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.99 (0.73, 1.36)</td>
</tr>
<tr>
<td></td>
<td>96-111</td>
<td>111-121</td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.95 (0.68, 1.32)</td>
</tr>
<tr>
<td></td>
<td>24-29</td>
<td>19-21</td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.44 (0.78, 2.68)</td>
</tr>
<tr>
<td>1960-1990 (20% MI, peak)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>249-271</td>
<td>269-291</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td></td>
<td>118-126</td>
<td>135-151</td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.91 (0.67, 1.25)</td>
</tr>
<tr>
<td></td>
<td>112-126</td>
<td>121-134</td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.02 (0.74, 1.40)</td>
</tr>
<tr>
<td></td>
<td>32-39</td>
<td>27-33</td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.29 (0.75, 2.24)</td>
</tr>
</tbody>
</table>

CCA: complete case analysis; CI: confidence interval; MI: multiple imputation; OR: odds ratio; PAH: polycyclic aromatic hydrocarbon

a) Combined over m=30 imputations. Sample size varies across imputed data sets.

b) Maximum annual exposure.
Table 3.4. Associations between residential vehicular traffic exposure and breast cancer risk, within strata of fruit/vegetable intake and menopausal status.

<table>
<thead>
<tr>
<th>Traffic Exposure Classifications</th>
<th>Cases</th>
<th>Controls</th>
<th>Age-adjusted OR (95% CI)</th>
<th>Cases</th>
<th>Controls</th>
<th>Age-adjusted OR (95% CI)</th>
<th>p for interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fruit/vegetable Intake</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995 (CCA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>404</td>
<td>431</td>
<td>1.00 (referent)</td>
<td>228</td>
<td>215</td>
<td>1.00 (referent)</td>
<td>0.01</td>
</tr>
<tr>
<td>50-75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>193</td>
<td>188</td>
<td>1.10 (0.86, 1.41)</td>
<td>98</td>
<td>159</td>
<td>0.58 (0.43, 0.80)</td>
<td></td>
</tr>
<tr>
<td>75-95&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>156</td>
<td>132</td>
<td>1.24 (0.94, 1.62)</td>
<td>99</td>
<td>120</td>
<td>0.78 (0.56, 1.08)</td>
<td></td>
</tr>
<tr>
<td>≥95&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>42</td>
<td>30</td>
<td>1.46 (0.89, 2.40)</td>
<td>28</td>
<td>29</td>
<td>0.92 (0.53, 1.60)</td>
<td></td>
</tr>
<tr>
<td>p for trend</td>
<td>0.04</td>
<td></td>
<td></td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960-1990 (20% MI)&lt;sup&gt;b,c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>169-182</td>
<td>174-200</td>
<td>1.00 (referent)</td>
<td>89-105</td>
<td>100-113</td>
<td>1.00 (referent)</td>
<td></td>
</tr>
<tr>
<td>50-75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>79-92</td>
<td>73-84</td>
<td>1.19 (0.80, 1.77)</td>
<td>35-45</td>
<td>59-70</td>
<td>0.67 (0.39, 1.13)</td>
<td></td>
</tr>
<tr>
<td>≥75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>79-89</td>
<td>56-64</td>
<td>1.43 (0.94, 2.16)</td>
<td>42-49</td>
<td>66-72</td>
<td>0.71 (0.43, 1.15)</td>
<td></td>
</tr>
<tr>
<td>p for trend</td>
<td>0.09</td>
<td></td>
<td></td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Menopausal Status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Premenopausal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995 (CCA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>190</td>
<td>229</td>
<td>1.00 (referent)</td>
<td>439</td>
<td>397</td>
<td>1.00 (referent)</td>
<td></td>
</tr>
<tr>
<td>50-75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>101</td>
<td>122</td>
<td>0.95 (0.68, 1.32)</td>
<td>193</td>
<td>222</td>
<td>0.79 (0.62, 1.00)</td>
<td></td>
</tr>
<tr>
<td>75-95&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>91</td>
<td>67</td>
<td>1.64 (1.13, 2.38)</td>
<td>166</td>
<td>186</td>
<td>0.80 (0.62, 1.02)</td>
<td></td>
</tr>
<tr>
<td>≥95&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>16</td>
<td>16</td>
<td>1.20 (0.58, 2.47)</td>
<td>51</td>
<td>43</td>
<td>1.06 (0.69, 1.63)</td>
<td></td>
</tr>
<tr>
<td>p for trend</td>
<td>0.04</td>
<td></td>
<td></td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960-1990 (20% MI)&lt;sup&gt;b,c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.50</td>
</tr>
<tr>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>49-66</td>
<td>56-73</td>
<td>1.00 (referent)</td>
<td>209-228</td>
<td>210-240</td>
<td>1.00 (referent)</td>
<td></td>
</tr>
<tr>
<td>50-75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>23-29</td>
<td>29-38</td>
<td>0.84 (0.42, 1.66)</td>
<td>95-108</td>
<td>101-114</td>
<td>1.00 (0.70, 1.42)</td>
<td></td>
</tr>
<tr>
<td>≥75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>23-29</td>
<td>19-26</td>
<td>1.31 (0.63, 2.71)</td>
<td>97-109</td>
<td>109-115</td>
<td>0.91 (0.64, 1.29)</td>
<td></td>
</tr>
<tr>
<td>p for trend</td>
<td>0.60</td>
<td></td>
<td></td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CCA: complete case analysis; CI: confidence interval; MI: multiple imputation; OR: odds ratio
a) 35+ vs. 0-34 servings/week (Gaudet et al. 2004).
b) Combined over m=30 imputations. Sample size varies across imputed data sets.
c) The upper quantiles were combined to avoid cell sizes of <10. The overall OR for the association between traffic exposure (≥75<sup>th</sup> vs. <50<sup>th</sup> percentile) and breast cancer is 1.02 (95% CI: 0.75, 1.39).
Table 3.5. Associations between residential vehicular traffic PAH exposure and the risk of breast cancer categorized by tumor hormone receptor status.

<table>
<thead>
<tr>
<th>Tumor Categories</th>
<th>Percentiles of Exposure</th>
<th>Cases</th>
<th>Controls</th>
<th>Age-adjusted OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER+/PR+ vs. other subtypes (ER-/PR+, ER+/PR-, or ER-/PR-)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER+ and PR+ breast cancer</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>281</td>
<td>659</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>105</td>
<td>353</td>
<td>0.69 (0.53, 0.90)</td>
</tr>
<tr>
<td></td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>94</td>
<td>261</td>
<td>0.82 (0.62, 1.08)</td>
</tr>
<tr>
<td></td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>23</td>
<td>61</td>
<td>0.86 (0.52, 1.41)</td>
</tr>
<tr>
<td>ER- or PR- breast cancer</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>192</td>
<td>659</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>70</td>
<td>353</td>
<td>0.68 (0.50, 0.91)</td>
</tr>
<tr>
<td></td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>61</td>
<td>261</td>
<td>0.78 (0.57, 1.08)</td>
</tr>
<tr>
<td></td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>21</td>
<td>61</td>
<td>1.15 (0.68, 1.94)</td>
</tr>
<tr>
<td>ER+ or PR+ breast cancer</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>376</td>
<td>659</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>134</td>
<td>353</td>
<td>0.66 (0.52, 0.83)</td>
</tr>
<tr>
<td></td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>124</td>
<td>261</td>
<td>0.81 (0.63, 1.03)</td>
</tr>
<tr>
<td></td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>29</td>
<td>61</td>
<td>0.80 (0.50, 1.27)</td>
</tr>
<tr>
<td>Ratio of ORs (top quantiles: 0.86/1.15): 0.74 (0.40, 1.38)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER-/PR- vs. other subtypes (ER+/PR-, ER-/PR+, or ER+/PR+)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER- and PR- breast cancer</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>97</td>
<td>659</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>41</td>
<td>353</td>
<td>0.79 (0.54, 1.16)</td>
</tr>
<tr>
<td></td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>31</td>
<td>261</td>
<td>0.81 (0.52, 1.24)</td>
</tr>
<tr>
<td></td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>15</td>
<td>61</td>
<td>1.67 (0.91, 3.05)</td>
</tr>
<tr>
<td>ER+ or PR+ breast cancer</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>376</td>
<td>659</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>134</td>
<td>353</td>
<td>0.66 (0.52, 0.83)</td>
</tr>
<tr>
<td></td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>124</td>
<td>261</td>
<td>0.81 (0.63, 1.03)</td>
</tr>
<tr>
<td></td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>29</td>
<td>61</td>
<td>0.80 (0.50, 1.27)</td>
</tr>
<tr>
<td>Ratio of ORs (top quantiles: 1.67/0.80): 2.09 (1.08, 4.06)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CI: confidence interval; ER: estrogen receptor; OR: odds ratio from polytomous logistic regression; PAH: polycyclic aromatic hydrocarbon; PR: progesterone receptor
Spearman correlation coefficients for the associations between the traffic exposure estimates were stronger than Pearson correlations, likely because Pearson correlation is more sensitive to outliers. For example, the Pearson and Spearman coefficients comparing 1995 and 1960-1990 (20% MI) exposures were 0.41 and 0.76, respectively. Outlier exposure estimates in our study reflect truly increased exposures for these individuals due to proximity to heavily trafficked intersections (Beyea et al. 2006, 2013). Thus, Pearson coefficients may be informative when considering women with the highest exposure levels, but Spearman coefficients are probably more representative of the overall ranking for most participants.

We also examined the association between traffic exposure during 1995 and breast cancer risk among women residing in their current home for 15 or more years, and found that the association was similar among these long-term residents (as compared to among all participants).

In addition, we conducted a sensitivity analysis removing the top 2 and top 5 participants with the highest exposure levels from regression models (1995, 1980-1995, 1960-1990) in order to assess the influence of outliers on the reported effect estimates. For 1960-1990 exposure (20% single imputation), we varied the relative contribution of the pre-arrival surrogate to total exposure (Beyea et al. 2013), removed the contribution of intersections from the exposure model, and imputed missing exposures according to census place (i.e. city, town, or village; Beyea et al. 2013). Emissions at intersections contribute strongly to total traffic PAH exposure because of the increased acceleration/deceleration that takes place there (Beyea et al. 2006).

Removing participants with the highest exposures attenuated effect estimates to a small extent (data not shown). Halving or doubling the contribution of the pre-arrival surrogate to cumulative exposure or imputing according to census division did not appreciably alter our
results (data not shown). After removing the contribution of intersections from the exposure model, we observed null associations between traffic PAH exposure and breast cancer risk (even among women with the highest exposures) which is consistent with a previous LIBCSP report indicating the importance of intersection emissions to total exposure (Beyea et al. 2006); however, interactions with fruit/vegetable intake were still evident (data not shown).
Table 3.6. Correlations between selected residential vehicular traffic PAH exposure variables.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spearman correlations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>1</td>
<td>0.95</td>
<td>0.85</td>
<td>0.84</td>
<td>0.76</td>
</tr>
<tr>
<td>1980-1995 CCA</td>
<td></td>
<td>1</td>
<td>0.92</td>
<td>0.92</td>
<td>0.88</td>
</tr>
<tr>
<td>1960-1995 CCA</td>
<td></td>
<td></td>
<td>1</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1960-1990, CCA</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1960-1990 20% MIa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pearson correlations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>1</td>
<td>0.91</td>
<td>0.59</td>
<td>0.58</td>
<td>0.41</td>
</tr>
<tr>
<td>1980-1995 CCA</td>
<td></td>
<td>1</td>
<td>0.71</td>
<td>0.69</td>
<td>0.70</td>
</tr>
<tr>
<td>1960-1995 CCA</td>
<td></td>
<td></td>
<td>1</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1960-1990 CCA</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1960-1990 20% MIa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CCA: complete case analysis; MI: multiple imputation; PAH: polycyclic aromatic hydrocarbon

a) Combined over m=30 imputations; sample size varies across imputed data sets (see text for methods).
CHAPTER IV: POLYMORPHISMS IN DNA REPAIR GENES, TRAFFIC-RELATED POLYCYCLIC AROMATIC HYDROCARBON EXPOSURE, AND BREAST CANCER RISK

INTRODUCTION

Breast cancer is the most common malignancy among women in the United States (ACS 2011-2012). Polycyclic aromatic hydrocarbons (PAHs) are incomplete combustion by-products found in air pollution (Bostrom et al. 2002, IARC 2010). In particular, vehicular traffic is a major contributor to ambient PAH levels, especially near cities (Harrison et al. 1996, Nielsen 1996).

PAHs are confirmed human lung carcinogens and potent mammary carcinogens in animal models (IARC 2010). However, the association between PAH exposure and breast cancer in women is still under investigation (Gammon and Santella 2008). PAH-DNA adducts (Gammon et al. 2002c, Rundle et al. 2000) and air pollution exposure (Bonner et al. 2005, Mordukhovich et al. 2013 [submitted], Nie et al. 2007) have been associated with breast cancer in prior epidemiological studies, including in our study population (Gammon et al. 2002c, Mordukhovich et al. 2013 [submitted]).

PAHs damage DNA by forming bulky adducts (Braithwaite et al. 1998) and by inducing oxidative stress (IARC 2010). If unrepaired, PAH-induced DNA damage can lead to somatic mutations in tumor suppressor genes or proto-oncogenes, thereby contributing to carcinogenesis (Gammon and Santella 2008). Hence, genetic polymorphisms that alter DNA repair capacity may modify associations between PAH-related exposures and breast cancer risk (Berwick and Vineis 2000). Of the four major DNA repair pathways (Christmann et al. 2003), nucleotide excision repair (NER) and base excision repair (BER) are most likely to mend PAH-induced
DNA damage, since the NER pathway repairs bulky DNA adducts and the BER pathway repairs oxidative DNA damage (Goode et al. 2002, Braithwaite et al. 1998, Robertson et al. 2009).

The few studies conducted to date suggest that polymorphisms in NER and BER genes may interact with PAH-related exposures, such as cigarette smoke or PAH-DNA adducts, with respect to breast cancer risk (Crew et al. 2007, Mechanic et al. 2006, Shen et al. 2005, Terry et al. 2004). However, the impact of DNA repair polymorphisms on the relationship between air pollution and breast cancer has not been evaluated previously. Our study examines the association between residential vehicular traffic exposure and breast cancer risk within strata of NER and BER genotypes, using the resources of a population-based investigation (Gammon et al. 2002b). Findings of this first study to examine these associations, if confirmed, will increase understanding of breast cancer etiology, clarify mechanisms linking traffic PAHs to breast cancer risk, and identify women who may be particularly susceptible to the carcinogenic effects of traffic pollution on the breast.

MATERIALS AND METHODS

Our study used data from the case-control component of the Long Island Breast Cancer Study Project (LIBCSP), a population-based investigation conducted among adult women (ages 20-98) residing in Nassau and Suffolk counties in Long Island, NY (Gammon et al. 2002b). This study was approved by the institutional review boards of all participating institutions.

Study Population

Case participants were diagnosed with a first primary invasive or in situ breast cancer between July 1996 and August 1997, and were recruited via rapid case ascertainment through contact with the pathology departments of local hospitals. Control participants were women with no personal history of breast cancer. Controls under age 65 were identified using random digit
dialing (Waksberg 1978), and those aged 65 years or older were identified via Health Care Finance Administration rosters. Control participants were frequency matched based on the expected age distribution among the cases. In total, 1,508 cases and 1,556 controls (82.1% and 62.7% of eligible participants, respectively) completed the case-control study interview. Most participants were post-menopausal and white, which is consistent with the underlying racial distribution of the study counties at the time of data collection (Gammon et al. 2002b).

Previous studies using LIBCSP resources reported positive associations between several PAH-related exposures and overall breast cancer risk: elevated PAH-DNA adducts in mononuclear cells (Gammon et al. 2002c), long-term residential exposure to environmental tobacco smoke (Gammon et al. 2004b), increased consumption of grilled/smoked meat (Steck et al. 2007), and the top 5% of long-term residential vehicular traffic exposures, compared with exposures below the median (Mordukhovich et al. 2013 [submitted]).

Study Questionnaire

Trained interviewers administered a two-hour structured questionnaire in participants' homes (Gammon et al. 2002b). The questionnaire evaluated detailed information regarding a wide range of factors used in our analyses, including lifetime residential history, demographic characteristics, reproductive and medical history, and body size (Gammon et al. 2002b).

Traffic Exposure Assessment

Participants' historical residential traffic-related PAH exposures were reconstructed using a model which incorporated lifetime residential history in the study counties, historical U.S. vehicular PAH emissions and roadway-specific traffic patterns, local meteorological data, pollutant dispersion factors, and excess emissions at intersections (Beyea et al. 2005, 2006, 2008, 2013). In a validation study, the model accurately predicted residential soil PAH measurements.
as well as participants' levels of PAH-DNA adducts in mononuclear cells (Beyea et al. 2006).

Exposure estimates were reconstructed for the years 1960-1990 and 1995. Some participants moved to the study area after 1960, and others had periods of incomplete address data while residing in the study counties. Hence, an exposure surrogate was developed to estimate pre-arrival exposures, with variance added during multiple imputation (MI), and missing post-arrival exposures were calculated via MI based on participants' known residential addresses (Beyea et al. 2013, Rubin 1996). The percentage of imputation was restricted to ≤20% of total 1960-1990 exposure (Mordukhovich et al. 2013 [submitted]). For 1995 exposure, only women with complete, unimputed exposure information were included in regressions (denoted as a complete case analysis (CCA)).

Because our exposure model predicted benzo[a]pyrene levels as a surrogate for all traffic PAHs, exposure estimates were normalized to the mean exposure level for the year 1995 and presented as relative units of average annual exposure during 1995 (Beyea et al. 2006, 2013). Exposures in the year 1995 ranged between 0.02 and 31 relative units (mean=1.0), and cumulative exposures during the years 1960-1990 (20% MI) ranged between 16 and 5467 relative units of average annual exposure in 1995 (mean=212) (Mordukhovich et al. 2013 [submitted]). Most LIBCSP participants had lived in their current home for at least 15 years (Gammon et al. 2002b), and previously reported correlations between exposures in 1995 and 1960-1990 were strong (Mordukhovich et al. 2013 [submitted]).

Blood Sample Collection and DNA Extraction

Study interviewers, who were either nurses or certified phlebotomists, obtained a 40 mL non-fasting blood sample at the study visit from 73% of both cases and control participants (Gammon et al. 2002b). Blood samples were collected in EDTA-treated tubes and shipped
overnight to a laboratory in Columbia University, where they were aliquoted and stored in -80°C freezers using participants’ randomly assigned study identification numbers (Gammon et al. 2002b).

Mononuclear cells were separated from whole blood using Ficoll (Sigma Chemical Co., St. Louis, MO), washed twice with phosphate-buffered saline, and frozen until DNA isolation (Gammon et al. 2002b). DNA was extracted using phenol and chloroform isoamyl alcohol, followed by RNase and proteinase K treatment (Gammon et al. 2002b, Shen et al. 2005).

Genotyping

Our study evaluated single nucleotide polymorphisms (SNPs) in the following genes: excision repair cross-complementing group 1 (ERCC1, 8092C/A), 8-oxoguanine DNA glycosylase 1 (OGG1, Ser326Cys), x-ray repair cross complementing group 1 (XRCC1, Arg399Gln and Arg194Trp), and xeroderma pigmentosum groups A (XP A, -4A/G), D (XP D, Lys751Gln and Asp312Asn), F (XP F, Arg415Gln), and G (XP G, Asp1104His). These SNPs were previously found to be in Hardy-Weinberg equilibrium among LIBCSP controls (Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005, Terry et al. 2004). Genotyping was completed for 91-97% of women who donated a blood sample, depending on the polymorphism (n=2,050-2,177); missing genotype data was mainly due to insufficient DNA available for assays (Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005, Terry et al. 2004).

Several high-throughput genotyping methods are available for use in epidemiological studies, and the methods used in the LIBCSP have changed over time. SNPs in ERCC1, XRCC1, OGG1 and XP D (Lys751Gln) were genotyped using fluorescence polarization (FP) (Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005, Terry et al. 2004), which identifies the polymorphic base using a template-directed, dye-labeled dideoxynucleotide (Shen et al. 2008). SNPs in XP A
and XPD (Asp312Asn) were genotyped using the Taqman assay (Applied Biosystems, Foster City, CA) (Crew et al. 2007), which uses fluorogenic oligonucleotide reverse probes for allele recognition. SNPs in XPF and XPG were genotyped using Sequenom's high-throughput matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, chosen for its ability to multiplex (BioServe Biotechnologies, Laurel, MD) (Crew et al. 2007).

Quality control measures included negative and positive controls for genotype on each plate (Crew et al. 2007). In addition, duplicates for approximately 10% of DNA samples were selected for re-sampling. As previously reported, concordance between duplicate samples was high for all SNPs (κ>90%; Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005, Terry et al. 2004). Laboratory personnel were blinded to case-control status and duplicate sample status.

**Statistical Methods**

Our study examined associations between tertiles of traffic-related PAH exposure (years 1995 and 1960-1990) and breast cancer risk within strata of NER and BER polymorphisms. We evaluated PAH exposures for the year 1995 in order to maximize statistical power for gene-environment interaction analyses. We also evaluated 1960-1990 exposure estimates, as these may be a more accurate reflection of long-term exposure opportunity. We chose a categorical exposure variable because of the previously reported non-linear relationship between traffic PAHs and breast cancer risk in the LIBCSP (Mordukhovich et al. 2013 [submitted]).

We used unconditional logistic regression to estimate odds ratios (ORs) and 95% confidence intervals (CIs) (Hosmer and Lemeshow 1989). Trend tests (n=45) were conducted within stratified models, and likelihood ratio tests (n=18) were used to assess multiplicative interactions between traffic PAHs and SNPs (Rothman and Greenland 1998). When evaluating 1960-1990 exposures, effect estimates and confidence limits were combined across 30 imputed
data sets using Rubin's rules (Rubin 1996). Statistical analyses were conducted using SAS version 9.3 (SAS Institute, Cary, NC). No adjustment for multiple comparisons was made.

Potential confounders (educational attainment, household income, race, religion, parity, age at first birth, body mass index (kg/m²)) were identified by a thorough literature review and analysis of causal diagrams (Mordukhovich et al. 2013 [submitted], Shrier and Platt 2008). No covariates were retained using the 10% change in estimate criterion (Rothman and Greenland 1998) relative to age-adjusted models (Mordukhovich et al. 2013 [submitted]). Hence, regression models are adjusted only for the frequency matching factor, five-year age group.

Participants with heterozygote, variant homozygote, and homozygous major genotypes were considered separately in our study. We chose this approach because there was no a priori reason to combine women with heterozygote and variant homozygote genotypes into a single group, and because previous LIBCSP reports evaluating interactions between DNA repair polymorphisms and PAH-related exposures suggested differences between participants with heterozygote and variant homozygote genotypes that were not present when evaluating the main effects of these SNPs on breast cancer risk (Crew et al. 2007, Terry et al. 2004). However, participants with heterozygote and variant homozygote genotypes were combined when tertile-specific cell sizes comprised fewer than 10 cases and 10 controls (Peduzzi et al. 1996).

Finally, in a post-hoc analysis, we combined individual SNPs showing evidence of interactions with traffic PAHs. Using a previously reported method (Crew et al. 2007, Mechanic et al. 2006), we examined traffic-breast cancer associations stratified according to the number of 'high-risk' alleles (0-2, 3-4, ≥5) from the following SNPs: XPD Lys751Gln (Gln), XRCC1 Arg194Trp (Trp), XRCC1 Arg399Gln (Arg), and OGG1 Ser326Cys (Ser).
RESULTS

The number of LIBCSP participants with available exposure estimates varied according to the traffic PAH exposure definition (1995 vs. 1960-1990 estimates) and the imputed data set (some women were not included in all 30 imputed data sets because they exceeded the 20% limit on imputation percentage for certain imputation draws; Beyea et al. 2013, Mordukhovich et al. 2013 [submitted], Rubin 1996). Sample size also varied by the polymorphisms considered (\textit{XPD}, \textit{XPF}, \textit{XPG}, \textit{ERCC1}, \textit{OGG1}, and \textit{XRCC1}). Specifically, the study population size with SNP assays completed varied as follows: 1995 traffic estimates, 848-915 cases and 913-964 controls; 1960-1990 traffic estimates, 332-429 cases and 368-474 controls (Tables 1 and 2).

Among women with available genotyping data, ORs for the association between traffic exposure (highest vs. lowest tertile) and breast cancer risk were 1.01 (95% CI: 0.80, 1.26) and 1.04 (95% CI: 0.81, 1.33) for 1995 and 1960-1990 exposures, respectively. We present associations between traffic PAHs and breast cancer risk within strata of DNA repair SNPs in Tables 1 and 2.

We found no evidence of interactions between traffic PAHs and the following SNPs: \textit{XPA} -4A/G, \textit{XPF} Arg415Gln, \textit{XPG} Asp1104His, \textit{ERCC1} 8092C/A, and \textit{XPD} Asp312Asn. This lack of heterogeneity within genotype strata was evident regardless of whether we considered traffic PAHs estimated for 1995 (Table 1) or 1960-1990 (Table 2).

The association between vehicular traffic exposure in 1995 and breast cancer risk was of larger magnitude and greater precision among women with the homozygous variant genotype for the \textit{XPD} Lys751Gln polymorphism: OR (highest vs. lowest tertile) = 2.27 (95% CI: 1.22, 4.23; \(p\) for trend=0.01) (Table 1). Among women with the homozygous major or heterozygous genotypes, the corresponding ORs were 0.88 (95% CI: 0.61, 1.26) and 0.91 (95% CI: 0.65, 1.26),
respectively (p-interaction=0.03). Results for this SNP were similar, though less precise, when evaluating cumulative exposure for the years 1960-1990 (Table 2).

The relationship between 1960-1990 traffic exposure and breast cancer risk was also stronger among women with at least one variant allele for the \textit{XRCC1} Arg194Trp polymorphism (OR=3.04, 95% CI:0.97, 9.51; p for trend=0.049, p-interaction=0.30), among women with the homozygous major genotype for \textit{XRCC1} Arg399Gln (OR=1.88, 95% CI: 1.04, 3.41; p for trend=0.04, p-interaction=0.59), and among women with the homozygous major genotype for \textit{OGG1} Ser326Cys (OR=1.77, 95% CI: 1.09, 2.88; p for trend=0.02, p-interaction=0.52) (Table 2). The results were similar, but less precise and of smaller magnitude, when evaluating 1995 exposures (Table 1).

We examined the associations between traffic PAHs (highest vs. lowest tertile) and breast cancer risk stratified by the number of high-risk alleles in \textit{XPD} (Lys751Gln), \textit{XRCC1} (Arg194Trp), \textit{XRCC1} (Arg399Gln), and \textit{OGG1} (Ser326Cys). These SNPs were selected based on results presented in Tables 1 and 2. The ORs were strongest among women with 5 or more 'high-risk' variants (1995: OR = 3.22, 95% CI: 1.56, 6.66, p for trend = 0.002; 1960-1990: OR = 3.95, 95% CI: 1.24, 12.58, p for trend = 0.02), with no clear evidence of a positive association among those with fewer (0-2 or 3-4) high-risk variants (Table 3).

**DISCUSSION**

In the first breast cancer study to evaluate interactions between vehicular traffic and DNA repair polymorphisms, the estimated association between traffic exposure (upper vs. lower tertiles) and breast cancer risk was of two-fold greater magnitude among women with the homozygous variant genotype for the \textit{XPD} Lys751Gln polymorphism. This finding is consistent with the existing literature regarding \textit{XPD}, PAH exposures, and breast cancer risk (Crew et al.)
2007, Tang et al. 2002, Terry et al. 2004). In addition, the association between long-term traffic exposure and breast cancer was of three-fold greater magnitude among those with the homozygous major genotype for XRCC1 Arg399Gln, and of nearly two-fold greater magnitude among women with the homozygous major genotype for OGG1 Ser326Cys or with at least one variant allele for XRCC1 Arg194Trp. However, the associated confidence intervals were wide. If replicated, our findings may have public health significance due to widespread exposure to traffic emissions worldwide.

Air pollution, an important PAH source (Bonner et al. 2002, Harrison et al. 1996, Nielsen 1996), has been associated with breast cancer risk in other population studies (Bonner et al. 2005, Nie et al. 2007). The LIBCSP previously reported positive associations between traffic PAHs and breast cancer risk among biologically plausible subgroups of women, including those with very high long-term exposures, premenopausal breast cancer, hormone-receptor negative breast tumors, or low intake of fruits/vegetables (Mordukhovich et al. 2013 [submitted]).

PAHs damage DNA by causing bulky adducts (Braithwaite et al. 1998) and oxidative lesions (Goode et al. 2002), and the resulting damage can be repaired through NER and BER pathways (Braithwaite et al. 1998, Goode et al. 2002, Robertson et al. 2009, Wood 1997). NER is a complex process involving many proteins, including XPA, which recognizes DNA damage, and XPD, which unwinds DNA at the lesion site; ERCC1, XPF and XPG excise the damaged nucleotide fragment (Wood 1997). Proteins involved in BER include OGG1, which recognizes DNA damage and removes the damaged nucleotide, and XRCC1, a scaffolding protein (Robertson et al. 2009). Functional variants in the DNA repair genes encoding these proteins may modify associations between traffic PAHs and breast cancer risk (Berwick and Vineis 2000).

Studies evaluating the main effects of DNA repair genotypes on breast cancer risk report
inconsistent associations for most NER and BER polymorphisms (Choi et al. 2003, Crew et al. 2007, Mechanic et al. 2006). Variants in \textit{XPD} are more consistently related to breast cancer, including in several meta-analyses (Manuguerra et al. 2006, Qiu et al. 2010). The LIBCSP likewise reports modest positive associations between breast cancer risk and polymorphisms in \textit{XPD} (Crew et al. 2007, Terry et al. 2004), but not other NER or BER genes (Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005).

Our study found that the association between vehicular traffic exposure and breast cancer risk was of two times greater magnitude among women with the homozygous variant genotype for the \textit{XPD} Lys751Gln polymorphism. This is consistent with a previous report from our study population, which observed that the association between breast cancer and variant allele homozygosity for this SNP was limited to those with PAH-DNA adduct levels above the median and to current smokers (Terry et al. 2004). The Lys751Gln polymorphism is also linked to PAH-related DNA damage in epidemiological studies (Matullo et al. 2003, Palli et al. 2001), including increased levels of PAH-DNA adducts in breast tissue (Tang et al. 2002). Hence, the literature regarding PAH-related exposures, the \textit{XPD} Lys751Gln polymorphism, and breast cancer risk has been consistent to date (Gammon and Santella 2008).

In contrast, associations with other NER and BER variants have been inconsistently reported. For example, we observed that effect estimates for the relationship between traffic PAHs and breast cancer were 3-fold higher among women with the homozygous major genotype for the \textit{XRCC1} Arg399Gln polymorphism, consistent with a study which reported an inverse association between the variant allele for this SNP and PAH-DNA adducts in lung tissue (Zienolddiny et al. 2006). However, the LIBCSP investigators have previously reported positive associations between the variant \textit{XRCC1}-399Gln allele and breast cancer among women with
detectable PAH-DNA adducts and among never smokers (Shen et al. 2005).

We also report associations of nearly two-fold greater magnitude between traffic exposure and breast cancer risk among women with at least one variant allele for the \(XRCC1\) Arg194Trp polymorphism and among those with the homozygous major genotype for the \(OGG1\) Ser326Cys polymorphism. Previous studies found no evidence of interactions between these polymorphisms and active smoking (\(OGG1, XRCC1\)) or PAH-DNA adducts (\(XRCC1\)) (Shen et al. 2005, Rossner et al. 2006). Consistent with our results, epidemiologic studies reported positive associations between the variant \(XRCC1\) Trp194Arg SNP and micronucleus frequencies among coke oven workers, who are exposed to high levels of PAHs (Leng et al. 2005), as well as PAH-DNA adducts in lung tissue (Zienolddiny et al. 2006). However, only the variant \(OGG1\) Ser326Cys polymorphism has been linked to DNA damage endpoints, such as oxidative DNA lesions (Tarng et al. 2001), in several investigations (Aka et al. 2004, Bravard et al. 2009).

Finally, we report no evidence of interactions between traffic PAHs and SNPs in \(XPA\), \(XPF\), \(XPG\), \(ERCC1\), and \(XPD\) (Asp312Asn only). These findings are consistent with a previous LIBCSP report evaluating interactions between PAH-DNA adducts, smoking, and polymorphisms in \(XPA, XPF,\) and \(XPG\) with respect to breast cancer risk (Crew et al. 2007). This same study reported evidence of interactions between PAH-DNA adducts, but not active or passive smoking, and the \(ERCC1\) 8092C/A and \(XPD\) Asp312Asn SNPs (Crew et al. 2007).

These inconsistencies in the literature could be due in part to the use of different PAH exposure surrogates across studies: we evaluated air pollution, while other investigations focused on cigarette smoking and PAH-DNA adducts. This complicates interpretation of study results, as PAH sources vary with regard to a range of properties, including exposure levels and pathways, particle size, and co-pollutants (IARC 2010). Differing results across sources of PAH exposure
may also suggest a more complicated role for certain polymorphisms. For example, DNA repair polymorphisms could interact differently with lower- and higher-level PAH exposures. Also, variations in study methodology, exposure and outcome definitions, and participant characteristics, as well as possible chance findings, may underlie inconsistent findings.

Interpretation of our study results requires careful consideration of our study limitations as well as our strengths. Our study had limited power for evaluating lower level interactions, particularly when examining long-term traffic exposures, which increases the likelihood of both false positive and false negative findings (Rothman and Greenland 1998). We also did not have sufficient statistical power to evaluate interactions between DNA repair polymorphisms and high-level residential traffic exposures among the most highly exposed 1-5% of study subjects (Mordukhovich et al. 2013 [submitted]), due to sample size constraints within genetic subgroups. This may have attenuated or obscured some associations. In addition, our results were not adjusted for multiple comparisons. However, we used a targeted pathway approach in carefully selecting NER and BER polymorphisms that were likely to interact with traffic PAH exposures (Braithwaite et al. 1998, Goode et al. 2002).

Ours is the first study to evaluate the impact of variation in DNA repair genes on the association between breast cancer risk and air pollution, a ubiquitous and generalizable exposure source (Bostrom et al. 2002). It is one of only a small number of breast cancer studies to examine interactions between DNA repair variants and any PAH-related exposure (Crew et al. 2007, Mechanic et al. 2006, Shen et al. 2005, Terry et al. 2004). Thus, we believe our investigation is an important addition to the sparse literature for this topic. Furthermore, our results for the \textit{XPD} Lys751Gln SNP are consistent with the growing literature regarding this DNA repair gene, PAH exposures, and breast cancer risk, which increases confidence in our
findings (Crew et al. 2007, Tang et al. 2002, Terry et al. 2004). Other strengths of our study include the use of a sophisticated, validated model to reconstruct historical traffic exposures (Beyea et al. 2006, 2013), and extensive quality control procedures which were implemented for genotyping assays and questionnaire data (Crew et al. 2007, Gammon et al. 2002b).

CONCLUSIONS

Associations between vehicular traffic and breast cancer were 2- to 3-fold greater among women with polymorphisms in certain nucleotide excision repair and base excision repair genes, although confidence intervals were wide. This is the first breast cancer study to examine interactions between air pollution and DNA repair polymorphisms, and our findings require confirmation in larger studies. Future breast cancer studies should also evaluate interactions between traffic exposures and genetic variants in oxidative stress and carcinogen metabolizing pathways. Our findings, if confirmed, may help identify women who are particularly susceptible to the carcinogenic effects of traffic pollution on the breast, and may help clarify mechanisms linking traffic PAHs to breast cancer risk.
Table 4.1. Associations between traffic polycyclic aromatic hydrocarbon (PAH) exposure in 1995 and breast cancer, stratified by DNA repair genotype.

<table>
<thead>
<tr>
<th>DNA Repair Polymorphisms</th>
<th>Genotype&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Cases (n)</th>
<th>Controls (n)</th>
<th>Tertile of exposure&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Age-adjusted OR (95% CI)</th>
<th>p for trend</th>
<th>p&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>XPA -4A/G</strong></td>
<td>GG</td>
<td>128</td>
<td>132</td>
<td>1</td>
<td>1.0</td>
<td>0.81</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>134</td>
<td>141</td>
<td>2</td>
<td>0.96 (0.68, 1.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>146</td>
<td>153</td>
<td>3</td>
<td>0.96 (0.69, 1.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GA</td>
<td>153</td>
<td>140</td>
<td>1</td>
<td>1.0</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>114</td>
<td>157</td>
<td>2</td>
<td>0.65 (0.46, 0.91)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>142</td>
<td>120</td>
<td>3</td>
<td>1.06 (0.75, 1.48)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>31</td>
<td>35</td>
<td>1</td>
<td>1.0</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>29</td>
<td>50</td>
<td>2</td>
<td>0.66 (0.34, 1.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>32</td>
<td>30</td>
<td>3</td>
<td>1.08 (0.53, 2.19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ERCC1 8092C/A</strong></td>
<td>CC</td>
<td>154</td>
<td>164</td>
<td>1</td>
<td>1.0</td>
<td>0.49</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>139</td>
<td>204</td>
<td>2</td>
<td>0.69 (0.51, 0.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>165</td>
<td>150</td>
<td>3</td>
<td>1.12 (0.81, 1.53)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CA</td>
<td>140</td>
<td>127</td>
<td>1</td>
<td>1.0</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>116</td>
<td>126</td>
<td>2</td>
<td>0.83 (0.59, 1.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>127</td>
<td>129</td>
<td>3</td>
<td>0.88 (0.62, 1.24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>17</td>
<td>18</td>
<td>1</td>
<td>1.0</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>24</td>
<td>18</td>
<td>2</td>
<td>1.40 (0.57, 3.46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>24</td>
<td>22</td>
<td>3</td>
<td>1.12 (0.46, 2.74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>XPF Arg415Gln</strong></td>
<td>GG</td>
<td>250</td>
<td>250</td>
<td>1</td>
<td>1.0</td>
<td>0.86</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>225</td>
<td>280</td>
<td>2</td>
<td>0.78 (0.61, 1.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>248</td>
<td>244</td>
<td>3</td>
<td>0.98 (0.76, 1.26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GA or AA</td>
<td>48</td>
<td>51</td>
<td>1</td>
<td>1.0</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>43</td>
<td>60</td>
<td>2</td>
<td>0.75 (0.43, 1.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>52</td>
<td>44</td>
<td>3</td>
<td>1.23 (0.69, 2.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>XPG Asp1104His</strong></td>
<td>GG</td>
<td>189</td>
<td>167</td>
<td>1</td>
<td>1.0</td>
<td>0.46</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>133</td>
<td>186</td>
<td>2</td>
<td>0.62 (0.46, 0.84)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>155</td>
<td>147</td>
<td>3</td>
<td>0.91 (0.67, 1.24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GC</td>
<td>92</td>
<td>111</td>
<td>1</td>
<td>1.0</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>113</td>
<td>118</td>
<td>2</td>
<td>1.10 (0.75, 1.62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>109</td>
<td>120</td>
<td>3</td>
<td>1.01 (0.69, 1.49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>17</td>
<td>20</td>
<td>1</td>
<td>1.0</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>17</td>
<td>26</td>
<td>2</td>
<td>0.77 (0.32, 1.89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>23</td>
<td>18</td>
<td>3</td>
<td>1.52 (0.62, 3.71)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>XPD Lys751Gln</strong></td>
<td>AA</td>
<td>123</td>
<td>124</td>
<td>1</td>
<td>1.0</td>
<td>0.47</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td></td>
<td>97</td>
<td>140</td>
<td>2</td>
<td>0.68 (0.47, 0.97)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>112</td>
<td>124</td>
<td>3</td>
<td>0.88 (0.61, 1.26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AC</td>
<td>148</td>
<td>131</td>
<td>1</td>
<td>1.0</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>141</td>
<td>157</td>
<td>2</td>
<td>0.77 (0.56, 1.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>155</td>
<td>146</td>
<td>3</td>
<td>0.91 (0.65, 1.26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>36</td>
<td>53</td>
<td>1</td>
<td>1.0</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>40</td>
<td>51</td>
<td>2</td>
<td>1.15 (0.64, 2.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>49</td>
<td>31</td>
<td>3</td>
<td>2.27 (1.22, 4.23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>SNP</td>
<td>Genotype</td>
<td>n1</td>
<td>n2</td>
<td>p1</td>
<td>p2</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>----------</td>
<td>----</td>
<td>----</td>
<td>------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td><strong>XPD</strong></td>
<td>Asp312Asn</td>
<td>GG</td>
<td>123</td>
<td>141</td>
<td>1.0</td>
<td>0.55</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>109</td>
<td>152</td>
<td>2.0</td>
<td>0.80</td>
<td>(0.56, 1.13)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>133</td>
<td>131</td>
<td>3.0</td>
<td>1.11</td>
<td>(0.79, 1.57)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>147</td>
<td>120</td>
<td>1.0</td>
<td>0.70</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>128</td>
<td>147</td>
<td>2.0</td>
<td>0.69</td>
<td>(0.49, 0.96)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>145</td>
<td>138</td>
<td>3.0</td>
<td>0.83</td>
<td>(0.59, 1.16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>41</td>
<td>46</td>
<td>1.0</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>38</td>
<td>50</td>
<td>2.0</td>
<td>0.88</td>
<td>(0.48, 1.59)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>42</td>
<td>33</td>
<td>3.0</td>
<td>1.42</td>
<td>(0.76, 2.65)</td>
</tr>
<tr>
<td><strong>XRCC1</strong></td>
<td>Arg399Gln</td>
<td>GG</td>
<td>124</td>
<td>137</td>
<td>1.0</td>
<td>0.91</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>108</td>
<td>123</td>
<td>2.0</td>
<td>0.95</td>
<td>(0.66, 1.35)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>121</td>
<td>133</td>
<td>3.0</td>
<td>0.98</td>
<td>(0.69, 1.39)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>157</td>
<td>141</td>
<td>1.0</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>139</td>
<td>183</td>
<td>2.0</td>
<td>0.66</td>
<td>(0.48, 0.91)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>166</td>
<td>135</td>
<td>3.0</td>
<td>1.06</td>
<td>(0.76, 1.46)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td>31</td>
<td>1.0</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td>45</td>
<td>2.0</td>
<td>0.79</td>
<td>(0.40, 1.53)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>34</td>
<td>36</td>
<td>3.0</td>
<td>0.92</td>
<td>(0.46, 1.83)</td>
</tr>
<tr>
<td><strong>XRCC1</strong></td>
<td>Arg194Trp</td>
<td>CC</td>
<td>277</td>
<td>266</td>
<td>1.0</td>
<td>0.70</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>250</td>
<td>308</td>
<td>2.0</td>
<td>0.76</td>
<td>(0.60, 0.96)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>277</td>
<td>267</td>
<td>3.0</td>
<td>0.95</td>
<td>(0.75, 1.21)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td>42</td>
<td>1.0</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>43</td>
<td>2.0</td>
<td>0.89</td>
<td>(0.47, 1.71)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>43</td>
<td>36</td>
<td>3.0</td>
<td>1.49</td>
<td>(0.79, 2.81)</td>
</tr>
<tr>
<td><strong>OGG1</strong></td>
<td>Ser326Cys</td>
<td>CC</td>
<td>191</td>
<td>197</td>
<td>1.0</td>
<td>0.75</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>149</td>
<td>202</td>
<td>2.0</td>
<td>0.75</td>
<td>(0.56, 1.00)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>183</td>
<td>173</td>
<td>3.0</td>
<td>1.06</td>
<td>(0.79, 1.42)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>101</td>
<td>97</td>
<td>1.0</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>111</td>
<td>123</td>
<td>2.0</td>
<td>0.84</td>
<td>(0.58, 1.24)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>115</td>
<td>113</td>
<td>3.0</td>
<td>0.94</td>
<td>(0.64, 1.39)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td>12</td>
<td>1.0</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>19</td>
<td>2.0</td>
<td>0.72</td>
<td>(0.26, 2.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>14</td>
<td>3.0</td>
<td>0.92</td>
<td>(0.32, 2.66)</td>
</tr>
</tbody>
</table>

a) Women with heterozygote and homozygote variant genotypes were combined to avoid cell sizes of <10
b) Traffic PAH exposures were calculated as units of average annual exposure in 1995. Tertiles: 0-0.70, 0.71-1.04, and ≥1.05 relative units. Tertile-specific exposure values are not presented because only relative rather than absolute traffic PAH exposures were estimated (see text for methods).
c) p for interaction.
Table 4.2. Associations between traffic polycyclic aromatic hydrocarbon (PAH) exposure in 1960-1990 and breast cancer, stratified by DNA repair genotype.

<table>
<thead>
<tr>
<th>DNA Repair Polymorphisms</th>
<th>Genotype&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Cases (n)</th>
<th>Controls (n)</th>
<th>Tertile of exposure&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Age-adjusted OR (95% CI)</th>
<th>p for trend</th>
<th>p&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>XPA -4A/G</strong></td>
<td>GG</td>
<td>51-63&lt;sup&gt;d&lt;/sup&gt;</td>
<td>63-74</td>
<td>1</td>
<td>1.0</td>
<td>0.74</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>67-67</td>
<td>53-62</td>
<td>2</td>
<td>1.32 (0.78, 2.24)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49-58</td>
<td>54-59</td>
<td>3</td>
<td>1.08 (0.63, 1.87)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GA or AA</td>
<td>60-75</td>
<td>76-92</td>
<td>1</td>
<td>1.0</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>57-67</td>
<td>84-97</td>
<td>2</td>
<td>0.87 (0.52, 1.43)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>78-90</td>
<td>60-76</td>
<td>3</td>
<td>1.50 (0.91, 2.47)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ERCC1 8092C/A</strong></td>
<td>CC</td>
<td>57-68</td>
<td>75-88</td>
<td>1</td>
<td>1.0</td>
<td>0.47</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>63-72</td>
<td>72-86</td>
<td>2</td>
<td>1.11 (0.68, 1.82)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>60-72</td>
<td>64-73</td>
<td>3</td>
<td>1.21 (0.73, 2.01)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CA or AA</td>
<td>56-71</td>
<td>64-78</td>
<td>1</td>
<td>1.0</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>56-66</td>
<td>66-76</td>
<td>2</td>
<td>0.89 (0.53, 1.51)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>68-75</td>
<td>50-61</td>
<td>3</td>
<td>1.39 (0.82, 2.36)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>XPF Arg415Gln</strong></td>
<td>GG</td>
<td>93-109</td>
<td>109-130</td>
<td>1</td>
<td>1.0</td>
<td>0.15</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>100-112</td>
<td>121-135</td>
<td>2</td>
<td>0.98 (0.66, 1.45)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>105-118</td>
<td>85-101</td>
<td>3</td>
<td>1.37 (0.91, 2.06)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GA or AA</td>
<td>17-25</td>
<td>24-31</td>
<td>1</td>
<td>1.0</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13-19</td>
<td>17-23</td>
<td>2</td>
<td>1.04 (0.41, 2.66)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17-21</td>
<td>19-24</td>
<td>3</td>
<td>1.03 (0.42, 2.53)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>XPG Asp1104His</strong></td>
<td>GG</td>
<td>64-78</td>
<td>68-82</td>
<td>1</td>
<td>1.0</td>
<td>0.34</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>57-68</td>
<td>79-92</td>
<td>2</td>
<td>0.81 (0.49, 1.34)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>67-78</td>
<td>53-67</td>
<td>3</td>
<td>1.31 (0.78, 2.18)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GC or CC</td>
<td>43-56</td>
<td>60-74</td>
<td>1</td>
<td>1.0</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>52-59</td>
<td>55-64</td>
<td>2</td>
<td>1.22 (0.69, 2.16)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49-58</td>
<td>53-61</td>
<td>3</td>
<td>1.22 (0.69, 2.14)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>XPD Lys751Gln</strong></td>
<td>AA</td>
<td>38-49</td>
<td>52-65</td>
<td>1</td>
<td>1.0</td>
<td>0.57</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>51-58</td>
<td>55-64</td>
<td>2</td>
<td>1.24 (0.71, 2.18)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47-53</td>
<td>51-59</td>
<td>3</td>
<td>1.19 (0.66, 2.12)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AC</td>
<td>59-70</td>
<td>61-73</td>
<td>1</td>
<td>1.0</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50-58</td>
<td>63-74</td>
<td>2</td>
<td>0.79 (0.46, 1.35)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>60-69</td>
<td>48-56</td>
<td>3</td>
<td>1.27 (0.74, 2.16)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>10-16</td>
<td>23-29</td>
<td>1</td>
<td>1.0</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17-21</td>
<td>18-23</td>
<td>2</td>
<td>1.77 (0.66, 4.74)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19-24</td>
<td>16-21</td>
<td>3</td>
<td>2.24 (0.80, 6.25)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>XPD Asp312Asn</strong></td>
<td>GG</td>
<td>47-57</td>
<td>58-74</td>
<td>1</td>
<td>1.0</td>
<td>0.61</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>51-61</td>
<td>63-73</td>
<td>2</td>
<td>1.07 (0.62, 1.83)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49-57</td>
<td>52-62</td>
<td>3</td>
<td>1.16 (0.66, 2.05)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GA or AA</td>
<td>64-81</td>
<td>79-94</td>
<td>1</td>
<td>1.0</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>65-77</td>
<td>75-88</td>
<td>2</td>
<td>1.02 (0.63, 1.66)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>77-87</td>
<td>63-74</td>
<td>3</td>
<td>1.40 (0.87, 2.26)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>XRCC1 Arg399Gln</strong></td>
<td>GG</td>
<td>36-45</td>
<td>56-69</td>
<td>1</td>
<td>1.0</td>
<td>0.039</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>49-60</td>
<td>50-59</td>
<td>2</td>
<td>1.51 (0.85, 2.68)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>55-65</td>
<td>41-51</td>
<td>3</td>
<td>1.88 (1.04, 3.41)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GA</td>
<td>60-74</td>
<td>62-79</td>
<td>1</td>
<td>1.0</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td></td>
<td>55-64</td>
<td>70-88</td>
<td>2</td>
<td>0.83 (0.50, 1.40)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Genotype: GG, GA, or AA

<sup>b</sup> Tertile of exposure: 1, 2, or 3

<sup>c</sup> p for trend

<sup>d</sup> p<sup>c</sup>
<table>
<thead>
<tr>
<th>Genotype</th>
<th>XRCC1 Arg194Trp</th>
<th>OGG1 Ser326Cys</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
<td>CC</td>
</tr>
<tr>
<td></td>
<td>CT or TT</td>
<td>GC or GG</td>
</tr>
<tr>
<td>AA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12-19</td>
<td>1.00 (0.61, 1.72)</td>
<td>1.00 (0.61, 1.72)</td>
</tr>
<tr>
<td>11-16</td>
<td>0.74 (0.25, 2.18)</td>
<td>1.17 (0.79, 1.73)</td>
</tr>
<tr>
<td>11-17</td>
<td>1.17 (0.79, 1.73)</td>
<td>0.87 (0.49, 1.54)</td>
</tr>
<tr>
<td>CC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>102-121</td>
<td>1.00 (0.44, 0.30)</td>
<td>3.04 (0.97, 9.51)</td>
</tr>
<tr>
<td>104-116</td>
<td>0.99 (0.68, 1.46)</td>
<td>0.87 (0.49, 1.54)</td>
</tr>
<tr>
<td>110-121</td>
<td>1.17 (0.79, 1.73)</td>
<td>0.84 (0.48, 1.47)</td>
</tr>
<tr>
<td>CT or TT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-14</td>
<td>1.00 (0.44, 0.30)</td>
<td>3.04 (0.97, 9.51)</td>
</tr>
<tr>
<td>16-20</td>
<td>1.61 (0.54, 4.76)</td>
<td>0.87 (0.49, 1.54)</td>
</tr>
<tr>
<td>19-23</td>
<td>3.04 (0.97, 9.51)</td>
<td>0.84 (0.48, 1.47)</td>
</tr>
<tr>
<td>GC or GG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46-57</td>
<td>1.00 (0.44, 0.30)</td>
<td>3.04 (0.97, 9.51)</td>
</tr>
<tr>
<td>50-60</td>
<td>0.87 (0.49, 1.54)</td>
<td>0.84 (0.48, 1.47)</td>
</tr>
<tr>
<td>49-56</td>
<td>0.87 (0.49, 1.54)</td>
<td>0.84 (0.48, 1.47)</td>
</tr>
</tbody>
</table>

a) Women with heterozygote and homozygote variant genotypes were combined to avoid cell sizes of <10
b) Traffic PAH exposures were calculated as units of average annual exposure in 1995. Tertiles: 0-144.50, 144.51-220.35, and ≥220.36 relative units.
c) p for interaction.
d) Effect estimates and their associated confidence intervals were combined across 30 imputed data sets; sample sizes vary across data sets because some women exceeded the 20% limit on imputation percentage for certain imputation draws.
Table 4.3. Associations between traffic PAH exposure and breast cancer, stratified by combined DNA repair genotypes in *XPD*, *XRCC1*, and *OGG1*.

<table>
<thead>
<tr>
<th>Number of High-Risk Alleles</th>
<th>1995 (CCA)</th>
<th>1960-1990 (≤20% MI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases (n)</td>
<td>Controls (n)</td>
</tr>
<tr>
<td>0-2</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>27</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td><em>p</em> for trend</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>106</td>
<td>96</td>
</tr>
<tr>
<td>84</td>
<td>104</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td><em>p</em> for trend</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-8</td>
<td>24</td>
<td>44</td>
</tr>
<tr>
<td>24</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td><em>p</em> for trend</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CCA: complete case analysis; MI: multiple imputation; PAH: polycyclic aromatic hydrocarbon (see text for methods).

a) Alleles were combined from the following genes: *XPD* Lys751Gln (Gln), *XRCC1* Arg194Trp (Arg), *XRCC1* Arg399Gln (Arg), and *OGG1* Ser326Cys (Ser).
b) Traffic PAH exposures were calculated as units of average annual exposure in 1995. Tertiles (1995): 0-0.70, 0.71-1.04, and ≥1.05 relative units. Tertiles (1960-1990): 0-144.50, 144.51-220.35, and ≥220.36 relative units.
c) Effect estimates and their associated confidence intervals were combined across 30 imputed data sets; sample sizes vary across data sets because some women exceeded the 20% limit on imputation percentage for certain imputation draws.
CHAPTER V: DISCUSSION

REVIEW OF STUDY AIMS

The primary aims of this dissertation were to evaluate associations between historical residential exposure to vehicular traffic-related polycyclic aromatic hydrocarbons (PAHs) and breast cancer risk, overall and within strata of nucleotide excision repair (NER) and base excision repair (BER) polymorphisms, using the resources of a population-based case-control study (the Long Island Breast Cancer Study Project (LIBCSP); Gammon et al. 2002b). Secondary aims of this study were to assess interactions between traffic PAHs and fruit and vegetable intake, as well as menopausal status, with respect to breast cancer risk, and to evaluate associations between traffic PAH exposure and breast cancer considered as subtypes of tumor hormone receptor status and somatic $p53$ mutation status.

SUMMARY OF RESULTS

Positive associations were observed between higher-level residential exposure to vehicular traffic PAHs and breast cancer risk. In adjusted cubic spline figures, an increased breast cancer risk among women with the top 1% of vehicular traffic exposure levels was evident, and in logistic regression models that compared participants with the top 5% of traffic PAH exposures to those with exposures below the median, effect estimates for breast cancer were increased by 14% for recent 1995 traffic exposures and by 44% for cumulative exposures during 1960-1990. Associations between traffic exposure in 1995 and breast cancer risk were increased by 46% among women with low fruit/vegetable intake, and by 67% and 64% among those with hormone receptor-negative breast tumors or premenopausal breast cancer,
respectively. However, confidence intervals for all of these estimates were wide.

The association between traffic PAH exposure (upper vs. lower tertiles) and breast cancer risk was also increased more than two-fold among women with the *XPD* Lys751Gln homozygous variant genotype. The association between cumulative 1960-1990 traffic exposure and breast cancer risk was three times higher for those with homozygous wild-type genotypes for *XRCC1* Arg399Gln, and nearly two times higher for women with homozygous wild-type genotypes for the *OGG1* Ser326Cys polymorphism or with at least one variant allele for the *XRCC1* Arg194Trp polymorphism. No interactions were observed between recent or long-term traffic PAH exposure and the following DNA repair polymorphisms: *XPA* -4A/G, *XPF* Arg415Gln, *XPG* Asp1104His, *ERCC1* 8092C/A, and *XPD* Asp312Asn.

**COMPARISON TO PREVIOUS INVESTIGATIONS**

Previous studies have reported positive associations between residential air pollution exposure and breast cancer (Bonner et al. 2005, Crouse et al. 2010, Lewis-Michl et al. 1996, Nie et al. 2007, Raaschou-Nielsen et al. 2011). In a breast cancer study conducted in Buffalo NY, effect estimates were increased over two-fold when comparing the highest and lowest quartiles of overall and traffic-related particulate pollution exposure in an area with high industrial emissions (Bonner et al. 2005, Nie et al. 2007). More modestly elevated estimates (<50%) were observed by other investigators when examining continuous traffic-related nitrogen oxide levels (Crouse et al. 2010, Raaschou-Nielsen et al. 2011) or high residential traffic density (Lewis-Michl et al. 1996).

Some of these previous studies relied on simple traffic density data (Lewis-Michl et al. 1996) or sparse monitors (Bonner et al. 2005) to generate exposure estimates, or lacked data on the carcinogenic particulate portion of air pollution (Crouse et al. 2010, Raaschou-Nielsen et al.
Only two studies used validated exposure assessment methods (Nie et al. 2007, Raaschou-Nielsen et al. 2011), and most studies focused on short-term exposure surrogates. For the study reported here, up to 35 years of individualized traffic-related particulate PAH exposure estimates were generated using a comprehensive exposure model, which was validated against extensive LIBCSP field data (Beyea et al. 2006). The consistency of positive associations across studies is encouraging given differences in air pollution exposure assessment, exposure levels and timing, and participant characteristics between study populations.

In our study, the association between residential traffic exposure and breast cancer risk was stronger among women who consumed low levels of fruits and vegetables. This is consistent with a previous LIBCSP report regarding breast cancer and grilled/smoked meat intake – another major PAH source (Steck et al. 2007). Fruit and vegetable intake has also been negatively associated with PAH-DNA adduct levels in the LIBCSP (Shantakumar et al. 2005) and in an epidemiologic study conducted in Europe (Palli et al. 2000). This is the first population-based breast cancer study to evaluate interactions between air pollution and dietary intake, and these findings require confirmation.

Traffic-breast cancer associations were also stronger among premenopausal women in the study reported here, consistent with previous LIBCSP reports regarding active and passive smoking (Gammon et al. 2004b) and PAH-DNA adduct levels (Gammon et al. 2002c). Similarly, a recent systematic review concluded that associations between environmental tobacco smoke and premenopausal, but not postmenopausal, breast cancer were “consistent with causality” (Johnson et al. 2011). Stronger associations were also noted between traffic PAH exposure and ER/PR-negative breast tumors, which are overrepresented among premenopausal women (Chen and Colditz 2007). Previous studies conducted in an area with high industrial
emissions reported positive associations between air pollution and breast cancer risk that varied inconsistently by menopausal status, depending on the exposure definition used, and did not vary by tumor hormone-receptor subtype (Bonner et al. 2005, Nie et al. 2007).

In this LIBCSP study, the association between traffic exposure and breast cancer risk was twice as strong among women with the homozygous variant genotype for the $XPD$ Lys751Gln polymorphism. This is consistent with a previous LIBCSP report, which observed that the association between breast cancer and variant allele homozygosity for this SNP was limited to those with PAH-DNA adduct levels above the median and to current smokers (Terry et al. 2004). The Lys751Gln polymorphism is also linked to PAH-related DNA damage in epidemiological studies (Matullo et al. 2003, Palli et al. 2001), including increased levels of PAH-DNA adducts in breast tissue (Tang et al. 2002). Hence, the literature regarding PAH-related exposures, the $XPD$, Lys751Gln polymorphism and breast cancer risk has been consistent to date.

The literature regarding breast cancer risk, PAH exposures, and the other variants examined in this study has been less consistent. For example, we observed a 3-fold higher effect estimate for the association between long-term traffic exposure and breast cancer risk among women with the homozygous wild-type genotype for $XRCC1$ Arg399Gln, consistent with a study which reported an inverse association between the variant allele for this SNP and PAH-DNA adducts in lung tissue (Zienolddiny et al. 2006). However, the LIBCSP investigators previously reported positive associations between the variant $XRCC1$-399Gln allele and breast cancer risk among women with detectable PAH-DNA adducts and among never smokers (Shen et al. 2005).

Associations between long-term traffic exposure and breast cancer risk were nearly two-fold higher among women with at least one variant allele for the $XRCC1$ Arg194Trp polymorphism and among those with the homozygous wild-type genotype for the $OGG1$
Ser326Cys polymorphism. Previous studies found no evidence of interactions between these polymorphisms and active smoking (OGG1, XRCC1) or PAH-DNA adducts (XRCC1) (Shen et al. 2005, Rossner et al. 2006). Consistent with our results, studies have reported positive associations between the variant XRCC1 Trp194Arg SNP and micronucleus frequencies among coke oven workers, who are exposed to high levels of PAHs (Leng et al. 2005), as well as PAH-DNA adducts in lung tissue (Zienolddiny et al. 2006). However, only the variant OGG1 Ser326Cys polymorphism has been linked to DNA damage, such as oxidative DNA lesions (Tarng et al. 2001), in several investigations (Aka et al. 2004, Bravard et al. 2009).

Finally, this study reported no evidence of interactions between traffic PAHs and SNPs in the following DNA repair genes: XPA, XPF, XPG, ERCC1, and XPD (Asp312Asn only). This is consistent with a previous LIBCSP report evaluating interactions between PAH-DNA adducts, smoking, and polymorphisms in XPA, XPF, and XPG (Crew et al. 2007). This latter study reported evidence of interactions between PAH-DNA adducts, but not active or passive smoking, and the ERCC1 8092C/A and XPD Asp312Asn polymorphisms (Crew et al. 2007).

Inconsistencies in the literature regarding PAH-related exposures, most DNA repair genes, and breast cancer risk could be due in part to the use of different PAH exposure surrogates across investigations. The current study evaluated air pollution, while other investigations focused on cigarette smoking and PAH-DNA adducts. This complicates interpretation of study results, as PAH sources vary with regard to a range of properties, including co-pollutants and typical exposure levels (IARC 2010). Differing results across sources of PAH exposure may also suggest a more complicated role for certain polymorphisms. For example, DNA repair polymorphisms could interact differently with lower- and higher-level PAH exposures (Dr. Regina Santella, personal communication 2013. In addition, variations in study methodology,
exposure and outcome definitions, and participant characteristics, as well as possible chance findings, may underlie inconsistent findings.

**BIOLOGICAL PLAUSIBILITY AND MECHANISMS**

Several mechanisms may underlie the potential association between traffic-related PAH exposure and breast cancer risk. PAHs are lipophilic and are stored in adipose tissue, including in the breast (Morris and Seifter 1992). When exposure levels are high or detoxification is insufficient, PAHs bind to DNA, forming PAH-DNA adducts, including in breast tissue (Gammon and Santella 2008, Santella 1999). PAHs have inflammatory (Jeng et al. 2010), pro-oxidant, and genotoxic (Farmer et al. 2003) properties, and induce oxidative DNA damage (Farmer et al. 2003). If uncorrected, PAH-induced DNA damage can cause somatic mutations in tumor suppressor genes or proto-oncogenes, which could in turn contribute to carcinogenesis (Gammon and Santella 2008). Several classes of biochemical pathways repair DNA damage. The NER pathway generally repairs DNA damage from bulky adducts, while the BER pathway generally repairs oxidative DNA damage (Goode et al. 2002, Braithwaite et al. 1998, Robertson et al. 2009). Hence, both DNA repair pathways could modify the effects of PAH exposures on breast cancer risk (Berwick and Vineis 2000).

In the study reported here, associations between traffic PAHs and breast cancer were stronger among women with low fruit/vegetable intake. This is consistent with the experimental evidence demonstrating that fruit and vegetable components have antioxidant and chemopreventive properties and are known to decrease the carcinogenic effects of PAH exposure in animals (Jin et al. 2006, Wattenberg and Loub 1978).
STUDY LIMITATIONS

Exposure Model

Traffic exposures were reconstructed for the years spanning 1960 and 1995. However, some participants moved to the study area after 1960, and others provided an incomplete address history while residing in the study area. Hence, limitations of our study include some missing address information, which was corrected in part through the use of multiple imputation (Beyea et al. 2013, Rubin 1996). Results for traffic-breast cancer associations were robust to varying the percentage of imputation (0-30%; see Chapter III).

Our model did not generate non-residential traffic exposure estimates, such as at work or while driving. A nation-wide study reported that, on average, American men and women spend almost 70% of their day inside (64.97%) or just outside (2.50%) of their residences, both overall and in a large city (Chicago, IL) (Leech et al. 2002). A Canadian study found similar results among women only (Nethery et al. 2009). In addition, the effect of time spent away from home may also be less than expected due to the large exposure component at night arising from pollutant drift from rush-hour traffic (Beyea et al. 2005).

The purpose of the exposure model was to estimate vehicular traffic exposures. Therefore, the model did not incorporate information on non-traffic sources of ambient PAHs, such as cooking, heating, leaf burning, or tobacco smoke (Beyea et al. 2005, 2006, 2013). Traffic is often the largest ambient source of PAHs near urban areas (Fromme et al. 2004, Nielsen 1996) and is more likely to release smaller and more inflammatory particles than some of the other ambient sources (Kocbach et al. 2006).

Benzo[a]pyrene was modeled as a surrogate for all traffic PAHs. Traffic pollution contains many chemicals, including many PAHs and other confirmed carcinogens (e.g.
benzene; Boström et al. 2002). It is impossible to ascribe health effects to a single traffic pollutant based on an observational study. Nevertheless, benzo[a]pyrene is one of the most carcinogenic PAHs and is representative of overall PAH exposure from vehicle exhaust (Boström et al. 2002). Particulate PAHs are among the most carcinogenic components of traffic exhaust, where they are found in high concentrations (Boothe and Shendell 2008). Furthermore, traffic pollutants do not have identical distributions (Beyea et al. 2006). In the LIBCSP, exposures were modeled using PAH-specific emissions data and dispersion parameters, and were validated and calibrated against soil PAH levels (Beyea et al. 2006). As an illustrative example, the LIBCSP traffic exposure model was altered (i.e. all depletion parameters were removed from the model) in order to predict carbon monoxide levels for a validation exercise.

Other Considerations

The study reported here had limited power for evaluating lower level interactions, particularly when examining long-term traffic exposures, where data was sparse. A smaller sample size increases the likelihood of both false positive and false negative findings (Rothman and Greenland 1998). This study also lacked sufficient statistical power to evaluate interactions between DNA repair polymorphisms and high-level residential traffic exposures among the most highly exposed 1-5% of study subjects (Chapter III), due to sample size constraints within genetic subgroups. This may have attenuated or obscured some associations. Interaction results were not adjusted for multiple comparisons. However, this study took a targeted pathway approach in selecting polymorphisms and dietary intake variables that were likely to interact with traffic PAHs, and is therefore much less likely to generate false-positive findings than if using an exploratory approach (Thompson 1994).

This study may be subject to potential limitations standard to epidemiological
investigations, such as laboratory errors with respect to genotyping or tumor classification, inaccurate reporting or recording of potentially confounding or modifying factors, and differential misclassification with respect to case-control status (Rothman and Greenland 1998). However, the LIBCSP has employed extensive quality control measures with regard to genotyping, tumor assays, and questionnaire data collection and storage procedures (described in the following section; Crew et al. 2007, Gammon et al. 2002b). These measures greatly reduce the probability of these potential study limitations (Gammon et al. 2002b).

Finally, the LIBCSP study population is fairly homogenous with respect to both socioeconomic status and race (Gammon et al. 2002b). This is both a limitation and a strength: participant homogeneity decreases the likelihood of residual confounding (Rothman and Greenland 1998), but reduces the generalizability of study findings.

STUDY STRENGTHS

**Exposure Model**

The study reported here used a comprehensive, individualized traffic exposure metric spanning a large geographic area and accounting for participants' residential histories for up to a 35 year period (Beyea et al. 2005, 2006). Vehicular traffic exposure estimates were derived using a sophisticated model that incorporated information on a wide range of factors including historical vehicular emissions and local transportation patterns, local meteorologic data, pollutant dispersion, deposition and decay factors, excess emissions from acceleration and deceleration at intersections, cruise and cold-engine emissions, building penetration factors, and background ambient PAHs from distant traffic sources (Beyea et al. 2006). The model was validated against extensive field data, (Beyea et al. 2006). Specifically, it was shown to accurately predict residential soil PAH levels, PAH-DNA adducts in mononuclear cells, and carbon monoxide
concentrations at an EPA monitoring site (Beyea et al. 2006). Participants were exposed to a wide range of residential traffic-related PAH levels, due to the gradient of urbanization across the study area (Beyea et al. 2006). Finally, breast cancer develops over many years (Clark et al. 1997), so a long modeled exposure period may be ideal for assessing relationships with breast cancer. A woman's traffic exposure level can vary greatly from year to year due to changes in emissions standards or moving to a new residence (Beyea et al. 2005, 2013).

Other Considerations

The LIBCSP utilized extensive quality control procedures for both questionnaire data and the laboratory component of the investigation. Specifically, approximately 10% of DNA samples were duplicated, distributed randomly on genotyping plates, and re-genotyped, and positive and negative controls were also included on each genotyping plate (Crew et al. 2007, Rossner et al. 2006, Shen et al. 2005, Terry et al. 2004). Responses from the case-control interview were checked using data range and logic checks, and a random sample of 20% of participants were briefly re-interviewed (Gammon et al. 2002b). Identical data collection and storage strategies were used for cases and controls, and laboratory personnel were blinded to both case-control status and duplicate status of participants' blood samples.

Additional strengths of this study include access to the resources of a large population-based investigation with extensive information regarding a wide range of factors, including demographic and lifestyle factors, active and passive cigarette smoking history, reproductive history, dietary intake, tumor markers and subtypes, and genetic polymorphisms in several PAH-related pathways (Gammon et al. 2002b). This facilitates effectively controlling for potential confounders, assessing effect modification, and evaluating tumor subtype-specific risk factor associations. The LIBCSP recruited incident rather than prevalent cases, which increases the
likelihood of identifying risk factors for breast cancer occurrence rather than survival (Gammon et al. 2002b, Rothman and Greenland 1998).

This is the first study to evaluate the impact of DNA repair polymorphisms or dietary intake on the association between air pollution exposure and breast cancer risk, and one of a small number of studies to examine interactions between DNA repair variants and any PAH-related exposure with respect to breast cancer risk (Crew et al. 2007, Mechanic et al. 2006, Shen et al. 2005, Terry et al. 2004). Hence, this analysis is an important addition to the sparse literature on this subject. Our results for XPD are consistent with the growing literature regarding this DNA repair gene, PAH exposures, and breast cancer risk (Crew et al. 2007, Pabalan et al. 2010, Tang et al. 2002, Terry et al. 2004).

FUTURE DIRECTIONS, PUBLIC HEALTH IMPLICATIONS, AND CONCLUSIONS

This study reports a positive association between breast cancer and high-level residential exposure to vehicular traffic PAHs reconstructed using a comprehensive, validated historical model. The association was most pronounced among women with low fruit/vegetable intake, hormone receptor negative tumors, premenopausal breast cancer, or certain NER or BER polymorphisms, though the associated confidence intervals were wide. Although U.S. traffic emissions are greatly reduced from the high levels of the 1960s and 1970s (Beyea et al. 2008), these findings may have public health significance given the relatively high incidence of breast cancer and ubiquitous exposure to traffic pollution worldwide, particularly in countries where vehicular traffic PAH emissions remain high (Han and Naeher 2006).

Future studies should examine associations between traffic exposure and breast cancer risk within larger or more highly exposed study populations, and in combination with other PAH sources (such as smoking or grilled/smoked foods). In addition to replicating the associations
evaluated in the current study, future investigations could examine interactions between traffic-related PAHs and intake of specific fruit/vegetable types or components, such as cruciferous vegetables, dietary antioxidants or phytonutrients, in order to evaluate whether particular fruits or vegetables are particularly beneficial for potential amelioration of breast cancer risk. Future studies could also evaluate interactions between air pollution and polymorphisms in biologically relevant genes, including in DNA repair, oxidative stress and carcinogen metabolizing pathways.

If confirmed, our results will contribute to increased understanding of breast cancer etiology, clarify mechanisms linking traffic to breast cancer and strengthen the biological plausibility of this relationship, and help identify women who may be especially susceptible to the carcinogenic effects of traffic pollution on the breast.
Figure A1.1: Metabolism of Polycyclic Aromatic Hydrocarbons in Relation to Breast Cancer Risk.

Source: Steck 2006  
(Adapted from Hecht 2002)
Table A2.1 Previously reported participant characteristics by case-control status

<table>
<thead>
<tr>
<th>Variables</th>
<th>Cases (n = 1,508)</th>
<th>Controls (n = 1,556)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean (SD)</strong></td>
<td><strong>N (%)</strong></td>
<td><strong>Mean (SD)</strong></td>
<td><strong>N (%)</strong></td>
</tr>
<tr>
<td>Age group at reference&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 35 years</td>
<td>39 (2.6)</td>
<td>45 (2.9)</td>
<td>0.0064</td>
</tr>
<tr>
<td>35-44 years</td>
<td>181 (12.0)</td>
<td>245 (15.8)</td>
<td></td>
</tr>
<tr>
<td>45-54 years</td>
<td>397 (26.3)</td>
<td>423 (27.2)</td>
<td></td>
</tr>
<tr>
<td>55-64 years</td>
<td>372 (24.7)</td>
<td>403 (25.9)</td>
<td></td>
</tr>
<tr>
<td>65-74 years</td>
<td>365 (24.2)</td>
<td>310 (19.9)</td>
<td></td>
</tr>
<tr>
<td>75-84 years</td>
<td>134 (8.9)</td>
<td>112 (7.2)</td>
<td></td>
</tr>
<tr>
<td>85+ years</td>
<td>20 (1.3)</td>
<td>18 (1.2)</td>
<td></td>
</tr>
<tr>
<td>Educational level&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>0.024</td>
</tr>
<tr>
<td>&lt; High school</td>
<td>183 (12.2)</td>
<td>150 (9.7)</td>
<td></td>
</tr>
<tr>
<td>High school graduate</td>
<td>538 (35.8)</td>
<td>526 (33.9)</td>
<td></td>
</tr>
<tr>
<td>Some college</td>
<td>360 (24.0)</td>
<td>415 (26.7)</td>
<td></td>
</tr>
<tr>
<td>College graduate</td>
<td>191 (12.7)</td>
<td>236 (15.2)</td>
<td></td>
</tr>
<tr>
<td>Post college</td>
<td>230 (15.3)</td>
<td>225 (14.5)</td>
<td></td>
</tr>
<tr>
<td>Income&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>0.033</td>
</tr>
<tr>
<td>&lt;$15,000</td>
<td>115 (7.7)</td>
<td>84 (5.4)</td>
<td></td>
</tr>
<tr>
<td>$15,000-$19,999</td>
<td>73 (4.9)</td>
<td>85 (5.5)</td>
<td></td>
</tr>
<tr>
<td>$20,000-$24,999</td>
<td>98 (6.5)</td>
<td>126 (8.1)</td>
<td></td>
</tr>
<tr>
<td>$25,000-$34,999</td>
<td>245 (16.3)</td>
<td>211 (13.6)</td>
<td></td>
</tr>
<tr>
<td>$35,000-$49,999</td>
<td>243 (16.2)</td>
<td>264 (17.0)</td>
<td></td>
</tr>
<tr>
<td>$50,000-$69,999</td>
<td>252 (16.8)</td>
<td>288 (18.5)</td>
<td></td>
</tr>
<tr>
<td>$70,000-$89,999</td>
<td>190 (12.6)</td>
<td>210 (13.5)</td>
<td></td>
</tr>
<tr>
<td>&lt;$90,000+</td>
<td>288 (19.2)</td>
<td>286 (18.4)</td>
<td></td>
</tr>
<tr>
<td>Race&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>0.073</td>
</tr>
<tr>
<td>White</td>
<td>1411 (93.8)</td>
<td>1429 (91.8)</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>69 (4.6)</td>
<td>85 (5.5)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>25 (1.7)</td>
<td>42 (2.7)</td>
<td></td>
</tr>
<tr>
<td>Menopausal status&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>0.29</td>
</tr>
<tr>
<td>Premenopausal</td>
<td>468 (31.7)</td>
<td>500 (33.5)</td>
<td></td>
</tr>
<tr>
<td>Postmenopausal</td>
<td>1010 (68.3)</td>
<td>993 (66.5)</td>
<td></td>
</tr>
<tr>
<td>Fruit/vegetable intake&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>0.046</td>
</tr>
<tr>
<td>0-34 servings/week</td>
<td>945 (63.9)</td>
<td>918 (60.4)</td>
<td></td>
</tr>
<tr>
<td>35+ servings/week</td>
<td>533 (36.1)</td>
<td>602 (39.6)</td>
<td></td>
</tr>
<tr>
<td>BMI at reference (kg/m&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>26.61 (5.7)</td>
<td>26.36 (5.8)</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>Cases</td>
<td>Control</td>
<td>p</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Parity</td>
<td>2.37 (1.6)</td>
<td>2.55 (1.7)</td>
<td>0.0021</td>
</tr>
<tr>
<td>Age at first birth(^d)</td>
<td>25.50 (4.7)</td>
<td>25.03 (4.6)</td>
<td>0.0093</td>
</tr>
</tbody>
</table>

BMI: body mass index; SD: Standard Deviation  

a) Cases and control participants were compared using Student's \(t\) test and chi-square analysis.  
b) Gammon et al. 2002b.  
c) Income was imputed for 11.5% of participants based on age, race, and education.  
Table A2.2. Associations between traffic exposure in the year 1995 and breast cancer, overall and among long-term residents $^a$.

<table>
<thead>
<tr>
<th>Exposure Years</th>
<th>Cases</th>
<th>Controls</th>
<th>Quantiles</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995 (overall)</td>
<td>645</td>
<td>659</td>
<td>&lt;50th percentile</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>299</td>
<td>353</td>
<td>50-&lt;75th percentile</td>
<td>0.87 (0.72, 1.04)</td>
</tr>
<tr>
<td></td>
<td>260</td>
<td>261</td>
<td>75-&lt;95th percentile</td>
<td>1.01 (0.82, 1.23)</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>61</td>
<td>≥95th percentile</td>
<td>1.14 (0.80, 1.64)</td>
</tr>
<tr>
<td>1995 (15+ year residents)</td>
<td>326</td>
<td>303</td>
<td>&lt;50th percentile</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>167</td>
<td>204</td>
<td>50-&lt;75th percentile</td>
<td>0.76 (0.59, 0.98)</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>146</td>
<td>75-&lt;95th percentile</td>
<td>0.80 (0.60, 1.06)</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>33</td>
<td>≥95th percentile</td>
<td>0.89 (0.53, 1.48)</td>
</tr>
<tr>
<td>1995 (overall)</td>
<td>645</td>
<td>659</td>
<td>&lt;50th percentile</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td></td>
<td>299</td>
<td>353</td>
<td>50-&lt;75th percentile</td>
<td>0.87 (0.72, 1.04)</td>
</tr>
<tr>
<td></td>
<td>312</td>
<td>313</td>
<td>75-&lt;99th percentile</td>
<td>1.00 (0.83, 1.22)</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>9</td>
<td>≥99th percentile</td>
<td>2.02 (0.90, 4.53)</td>
</tr>
<tr>
<td>1995 (15+ year residents)</td>
<td>326</td>
<td>303</td>
<td>&lt;50th percentile</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>167</td>
<td>204</td>
<td>50-&lt;75th percentile</td>
<td>0.76 (0.59, 0.98)</td>
</tr>
<tr>
<td></td>
<td>151</td>
<td>177</td>
<td>75-&lt;99th percentile</td>
<td>0.78 (0.59, 1.02)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2</td>
<td>≥99th percentile</td>
<td>4.39 (0.95, 20.30)</td>
</tr>
</tbody>
</table>

CI: confidence interval; OR: odds ratio

$^a$ Defined as women residing in their current home for 15 or more years.
Table A2.3. Associations between selected residential traffic exposure variables and breast cancer, removing participants with the highest exposure levels.

<table>
<thead>
<tr>
<th>Exposure Years</th>
<th>Cases</th>
<th>Controls</th>
<th>Quantiles</th>
<th>OR (95% CI)&lt;sup&gt;a&lt;/sup&gt; top 2 removed</th>
<th>OR (95% CI)&lt;sup&gt;a&lt;/sup&gt; top 5 removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995 (CCA)</td>
<td>645</td>
<td>659</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>299</td>
<td>353</td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.87 (0.72, 1.04)</td>
<td>0.87 (0.72, 1.04)</td>
</tr>
<tr>
<td></td>
<td>260</td>
<td>261</td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.01 (0.82, 1.23)</td>
<td>1.01 (0.82, 1.23)</td>
</tr>
<tr>
<td></td>
<td>68, 66</td>
<td>61, 60</td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt; (vs 1.14)</td>
<td>1.11 (0.77, 1.60)</td>
<td>1.10 (0.76, 1.59)</td>
</tr>
<tr>
<td>1980-1995 (CCA)</td>
<td>431</td>
<td>428</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>201</td>
<td>230</td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.86 (0.68, 1.08)</td>
<td>0.86 (0.68, 1.08)</td>
</tr>
<tr>
<td></td>
<td>171</td>
<td>173</td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.95 (0.74, 1.22)</td>
<td>0.95 (0.74, 1.22)</td>
</tr>
<tr>
<td></td>
<td>41, 38</td>
<td>42</td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt; (vs. 0.98)</td>
<td>0.93 (0.59, 1.46)</td>
<td>0.86 (0.54, 1.37)</td>
</tr>
<tr>
<td>1960-1990 (20% MI)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>262-287</td>
<td>289-320</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>122-139</td>
<td>136-155</td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.99 (0.73, 1.36)</td>
<td>0.99 (0.73, 1.36)</td>
</tr>
<tr>
<td></td>
<td>96-111</td>
<td>111-121</td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.95 (0.68, 1.32)</td>
<td>0.95 (0.68, 1.32)</td>
</tr>
<tr>
<td></td>
<td>22-27, 19-24</td>
<td>19-21, 19-21</td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt; (vs.1.44)</td>
<td>1.32 (0.70, 2.48)</td>
<td>1.20 (0.63, 2.28)</td>
</tr>
</tbody>
</table>

CCA: complete case analysis; CI: confidence interval; MI: multiple imputation; OR: odds ratio
a) Models are age-adjusted, as cases and control participants were frequency matched by age group.
b) Partially imputed datasets are combined over m = 30 imputations using Rubin's rules. Sample size varies by the imputed data set (see text for methods).
Table A2.4. Associations between traffic polycyclic aromatic hydrocarbon exposure and breast cancer, varying the relative contribution of the pre-arrival surrogate.

<table>
<thead>
<tr>
<th>Exposure Years</th>
<th>Cases</th>
<th>Controls</th>
<th>Quantiles</th>
<th>Age-adjusted OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases</td>
<td>Controls</td>
<td>Quantiles</td>
<td>Age-adjusted OR (95% CI)</td>
</tr>
<tr>
<td>1960-1990 (doubled)</td>
<td>254-279</td>
<td>279-305</td>
<td>&lt;50th percentile</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td></td>
<td>124-141</td>
<td>141-160</td>
<td>50-&lt;75th percentile</td>
<td>0.95 (0.69, 1.30)</td>
</tr>
<tr>
<td></td>
<td>96-121</td>
<td>112-126</td>
<td>75-&lt;95th percentile</td>
<td>0.95 (0.68, 1.33)</td>
</tr>
<tr>
<td></td>
<td>25-32</td>
<td>21-25</td>
<td>≥95th percentile</td>
<td>1.37 (0.75, 2.50)</td>
</tr>
<tr>
<td>1960-1990 (halved)</td>
<td>264-290</td>
<td>295-324</td>
<td>&lt;50th percentile</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td></td>
<td>120-139</td>
<td>132-148</td>
<td>50-&lt;75th percentile</td>
<td>1.02 (0.75, 1.40)</td>
</tr>
<tr>
<td></td>
<td>93-107</td>
<td>108-117</td>
<td>75-&lt;95th percentile</td>
<td>0.94 (0.67, 1.30)</td>
</tr>
<tr>
<td></td>
<td>24-28</td>
<td>19-21</td>
<td>≥95th percentile</td>
<td>1.46 (0.78, 2.71)</td>
</tr>
</tbody>
</table>

CI: confidence interval; MI: multiple imputation; OR: odds ratio

a) Up to 20% of dose was imputed for each exposure variable (single imputation; see text for methods).
Table A2.5. Associations between traffic PAH exposure and breast cancer, turning off the intersection component of the exposure model.

<table>
<thead>
<tr>
<th>Exposure Years $^a$</th>
<th>Cases</th>
<th>Controls</th>
<th>Quantiles</th>
<th>Age-adjusted OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960-1990</td>
<td>283</td>
<td>320</td>
<td>&lt;50$^{th}$ percentile</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td></td>
<td>133</td>
<td>122</td>
<td>50-&lt;75$^{th}$ percentile</td>
<td>1.22 (0.91, 1.64)</td>
</tr>
<tr>
<td></td>
<td>105</td>
<td>114</td>
<td>75-&lt;95$^{th}$ percentile</td>
<td>1.02 (0.75, 1.40)</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>23</td>
<td>≥95$^{th}$ percentile</td>
<td>0.90 (0.48, 1.70)</td>
</tr>
<tr>
<td>1960-1990</td>
<td>283</td>
<td>320</td>
<td>&lt;50$^{th}$ percentile</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td></td>
<td>133</td>
<td>122</td>
<td>50-&lt;75$^{th}$ percentile</td>
<td>1.22 (0.91, 1.64)</td>
</tr>
<tr>
<td></td>
<td>119</td>
<td>130</td>
<td>75-&lt;99$^{th}$ percentile</td>
<td>1.02 (0.75, 1.37)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>7</td>
<td>≥99$^{th}$ percentile</td>
<td>0.75 (0.23, 2.40)</td>
</tr>
</tbody>
</table>

CCA: complete case analysis; CI: confidence interval; MI: multiple imputation; OR: odds ratio; PAH: polycyclic aromatic hydrocarbon

$a$ Up to 20% of dose was imputed for each exposure variable (single imputation; see text for methods).
Table A2.6. Associations between traffic PAH exposure and breast cancer, with partial imputation conducted using imputation by census place.

<table>
<thead>
<tr>
<th>Exposure Years&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Cases</th>
<th>Controls</th>
<th>Quantiles</th>
<th>Age-adjusted OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960-1990</td>
<td>262</td>
<td>301</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td></td>
<td>133</td>
<td>139</td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>1.09 (0.81, 1.45)</td>
</tr>
<tr>
<td></td>
<td>110</td>
<td>115</td>
<td>75-&lt;95&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>1.07 (0.78, 1.46)</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>22</td>
<td>≥95&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>1.27 (0.70, 2.31)</td>
</tr>
<tr>
<td>1960-1990</td>
<td>262</td>
<td>301</td>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td></td>
<td>133</td>
<td>139</td>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>1.09 (0.81, 1.45)</td>
</tr>
<tr>
<td></td>
<td>125</td>
<td>135</td>
<td>75-&lt;99&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>1.03 (0.77, 1.39)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2</td>
<td>≥99&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>5.85 (1.26, 27.13)</td>
</tr>
</tbody>
</table>

CCA: complete case analysis; CI: confidence interval; MI: multiple imputation; OR: odds ratio; PAH: polycyclic aromatic hydrocarbon

a) Up to 20% of dose was imputed for each exposure variable (single imputation).
Table A2.7. Associations between selected residential traffic exposure variables and breast cancer categorized by tumor characteristics.

<table>
<thead>
<tr>
<th>Tumor Characteristics</th>
<th>Cases (1995 (CCA))</th>
<th>Controls (1995 (CCA))</th>
<th>Age-adjusted OR (95% CI)</th>
<th>Cases (20% MI)</th>
<th>Controls (20% MI)</th>
<th>Age-adjusted OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ER+/PR+ breast cancer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>659</td>
<td>281</td>
<td>1</td>
<td>289-320</td>
<td>125-144</td>
<td>1</td>
</tr>
<tr>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>353</td>
<td>105</td>
<td>0.69 (0.53, 0.90)</td>
<td>136-155</td>
<td>35-44</td>
<td>0.61 (0.39, 0.95)</td>
</tr>
<tr>
<td>≥75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>322</td>
<td>117</td>
<td>0.83 (0.64, 1.07)</td>
<td>132-140</td>
<td>46-62</td>
<td>0.80 (0.54, 1.20)</td>
</tr>
<tr>
<td><strong>ER+/PR- or ER-/PR+ or ER-/PR- breast cancer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>659</td>
<td>192</td>
<td>1</td>
<td>289-320</td>
<td>73-82</td>
<td>1</td>
</tr>
<tr>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>353</td>
<td>70</td>
<td>0.68 (0.50, 0.91)</td>
<td>136-155</td>
<td>27-34</td>
<td>0.82 (0.50, 1.35)</td>
</tr>
<tr>
<td>≥75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>322</td>
<td>82</td>
<td>0.85 (0.64, 1.14)</td>
<td>132-140</td>
<td>30-36</td>
<td>0.86 (0.53, 1.39)</td>
</tr>
<tr>
<td><strong>ER-/PR- breast cancer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>659</td>
<td>97</td>
<td>1</td>
<td>289-320</td>
<td>36-44</td>
<td>1</td>
</tr>
<tr>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>353</td>
<td>41</td>
<td>0.79 (0.54, 1.16)</td>
<td>136-155</td>
<td>12-16</td>
<td>0.75 (0.38, 1.48)</td>
</tr>
<tr>
<td>≥75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>322</td>
<td>46</td>
<td>0.97 (0.67, 1.41)</td>
<td>132-140</td>
<td>13-18</td>
<td>0.86 (0.45, 1.64)</td>
</tr>
<tr>
<td><strong>ER+/PR- or ER-/PR+ or ER+/PR+ breast cancer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>659</td>
<td>376</td>
<td>1</td>
<td>289-320</td>
<td>163-182</td>
<td>1</td>
</tr>
<tr>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>353</td>
<td>134</td>
<td>0.66 (0.52, 0.83)</td>
<td>136-155</td>
<td>49-62</td>
<td>0.67 (0.45, 0.99)</td>
</tr>
<tr>
<td>≥75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>322</td>
<td>153</td>
<td>0.80 (0.64, 1.01)</td>
<td>132-140</td>
<td>62-70</td>
<td>0.82 (0.57, 1.18)</td>
</tr>
<tr>
<td><strong>p53 mutation-positive breast cancer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>659</td>
<td>58</td>
<td>1</td>
<td>289-320</td>
<td>18-23</td>
<td>1</td>
</tr>
<tr>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>353</td>
<td>29</td>
<td>0.93 (0.58, 1.48)</td>
<td>136-155</td>
<td>6-13</td>
<td>0.89 (0.36, 2.21)</td>
</tr>
<tr>
<td>≥75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>322</td>
<td>23</td>
<td>0.79 (0.48, 1.31)</td>
<td>132-140</td>
<td>10-13</td>
<td>1.14 (0.52, 2.49)</td>
</tr>
<tr>
<td><strong>p53 mutation-negative breast cancer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>659</td>
<td>293</td>
<td>1</td>
<td>289-320</td>
<td>137-154</td>
<td>1</td>
</tr>
<tr>
<td>50-&lt;75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>353</td>
<td>142</td>
<td>0.90 (0.71, 1.14)</td>
<td>136-155</td>
<td>63-72</td>
<td>0.99 (0.68, 1.44)</td>
</tr>
<tr>
<td>≥75&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>322</td>
<td>172</td>
<td>1.16 (0.92, 1.47)</td>
<td>132-140</td>
<td>65-73</td>
<td>1.00 (0.69, 1.44)</td>
</tr>
</tbody>
</table>

CCA: complete case analysis; CI: confidence interval; ER: estrogen receptor; MI: multiple imputation; OR: odds ratio; PR: progesterone receptor

a) All partially imputed datasets are combined over m=30 imputations using Rubin's rules (see text for methods); sample size varies by imputation draw.
Table A2.8. Previously reported main effects of DNA repair polymorphisms on breast cancer risk.

<table>
<thead>
<tr>
<th>Polymorphisms</th>
<th>Genotypes</th>
<th>Cases, n (%)</th>
<th>Controls, n (%)</th>
<th>Age-adjusted OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>XPA</strong> -4A/G (rs1800975) a**</td>
<td>GG</td>
<td>488 (46.1)</td>
<td>488 (44.3)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>GA</td>
<td>466 (44.0)</td>
<td>477 (43.3)</td>
<td>0.97 (0.81, 1.17)</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>105 (9.9)</td>
<td>137 (12.4)</td>
<td>0.77 (0.58, 1.02)</td>
</tr>
<tr>
<td><strong>ERCC1</strong> 8092C/A (rs3212986) a**</td>
<td>CC</td>
<td>551 (52.1)</td>
<td>606 (54.9)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>CA</td>
<td>434 (41.1)</td>
<td>436 (39.5)</td>
<td>1.09 (0.92, 1.30)</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>72 (6.8)</td>
<td>62 (5.6)</td>
<td>1.29 (0.90, 1.85)</td>
</tr>
<tr>
<td><strong>XPD</strong> Lys751Gln (rs13181) b**</td>
<td>AA (Lys/Lys)</td>
<td>387 (36.8)</td>
<td>453 (41.1)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>AC (Lys/Gln)</td>
<td>513 (48.7)</td>
<td>498 (45.2)</td>
<td>1.22 (1.01, 1.46)</td>
</tr>
<tr>
<td></td>
<td>CC (Gln/Gln)</td>
<td>153 (14.5)</td>
<td>151 (13.7)</td>
<td>1.18 (0.91, 1.53)</td>
</tr>
<tr>
<td><strong>XPD</strong> Asp312Asn (rs1799793) a**</td>
<td>GG (Asp/Asp)</td>
<td>415 (40.2)</td>
<td>490 (45.2)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>GA (Asp/Asn)</td>
<td>478 (46.4)</td>
<td>454 (41.9)</td>
<td>1.25 (1.04, 1.50)</td>
</tr>
<tr>
<td></td>
<td>AA (Asn/Asn)</td>
<td>138 (13.4)</td>
<td>139 (12.8)</td>
<td>1.16 (0.89, 1.52)</td>
</tr>
<tr>
<td><strong>XPF</strong> Arg415Gln (rs1800067) a**</td>
<td>GG (Arg/Arg)</td>
<td>859 (84.4)</td>
<td>888 (83.4)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>GA (Arg/Gln)</td>
<td>156 (15.3)</td>
<td>167 (15.7)</td>
<td>0.99 (0.78, 1.26)</td>
</tr>
<tr>
<td></td>
<td>AA (Gln/Gln)</td>
<td>3 (0.3)</td>
<td>10 (0.9)</td>
<td>0.27 (0.07, 1.00)</td>
</tr>
<tr>
<td><strong>XPG</strong> Asp1104His (rs17655) a**</td>
<td>GG (Asp/Asp)</td>
<td>562 (56.3)</td>
<td>571 (54.3)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>GC (Asp/His)</td>
<td>371 (37.1)</td>
<td>409 (38.9)</td>
<td>0.94 (0.78, 1.13)</td>
</tr>
<tr>
<td></td>
<td>CC (His/His)</td>
<td>66 (6.6)</td>
<td>71 (6.8)</td>
<td>0.98 (0.69, 1.41)</td>
</tr>
<tr>
<td><strong>XRCC1</strong> Arg194Trp (rs1799782)</td>
<td>CC (Arg/Arg)</td>
<td>948 (88.0)</td>
<td>974 (87.3)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>CT (Arg/Trp)</td>
<td>125 (11.6)</td>
<td>134 (12.0)</td>
<td>0.96 (0.74, 1.24)</td>
</tr>
<tr>
<td></td>
<td>TT (Trp/Trp)</td>
<td>4 (0.4)</td>
<td>8 (0.7)</td>
<td>0.50 (0.15, 1.66)</td>
</tr>
<tr>
<td><strong>XRCC1</strong> Arg399Gln (rs25487) c**</td>
<td>GG (Arg/Arg)</td>
<td>412 (38.6)</td>
<td>444 (40.0)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>GA (Arg/Gln)</td>
<td>539 (50.5)</td>
<td>536 (48.3)</td>
<td>1.08 (0.90, 1.29)</td>
</tr>
<tr>
<td></td>
<td>AA (Gln/Gln)</td>
<td>116 (10.9)</td>
<td>130 (11.7)</td>
<td>0.97 (0.73, 1.29)</td>
</tr>
<tr>
<td><strong>OGG1</strong> Ser326Cys (rs1052133) d**</td>
<td>CC (Ser/Ser)</td>
<td>615 (59.1)</td>
<td>653 (59.7)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>GC (Ser/Cys)</td>
<td>375 (36.0)</td>
<td>385 (35.2)</td>
<td>1.04 (0.87, 1.24)</td>
</tr>
<tr>
<td></td>
<td>CC (Cys/Cys)</td>
<td>51 (4.9)</td>
<td>55 (5.0)</td>
<td>0.99 (0.66, 1.47)</td>
</tr>
</tbody>
</table>

OR: odds ratio; CI: confidence interval
a) Crew et al. 2007
b) Terry et al. 2004
c) Shen et al. 2005
d) Rossner et al. 2006


Bråbäck L, Forsberg B. 2009. Does traffic exhaust contribute to the development of asthma and allergic sensitization in children: findings from recent cohort studies. Environ Health 8:17


Caldecott KW. 2003. XRCC1 and DNA strand break repair. DNA Repair (Amst) 2(9):955-969.


Cavalieri EL, Rogan EG. 1998. Role of aromatic hydrocarbons in disclosing how


199


Friedenreich CM, Slimani N, Riboli E. 1992. Measurement of past diet: review of previous and


Gao WM, Romkes M, Siegfried JM, Luketich JD, Keohavong P. 2006. Polymorphisms in DNA


Glatt H, Harvey RG, Phillips DH, Hewer A, Grover PL. 1989. Influence of the alkyl substituent on mutagenicity and covalent DNA binding of bay region diol-epoxides of 7-methyl- and 7-


International Agency for Research on Cancer. 1989. Diesel and Gasoline Engine Exhausts and
Some Nitroarenes. Monographs on the Carcinogenic Risk of Chemicals to Humans. Lyon, France: IARC.


Kohno T, Shinmura K, Tosaka M, Tani M, Kim SR, Sugimura H, et al. 1998. Genetic polymorphisms and alternative splicing of the hOGG1 gene, that is involved in the repair of 8-
hydroxyguanine in damaged DNA. Oncogene 16(25):3219-3225.


polymorphisms have no measurable effect on nucleotide excision repair and basal transcription. DNA Repair (Amst) 6(9):1264-1270.


Manuguerra M, Saletta F, Karagas MR, Berwick M, Veglia F, Vineis P, et al. 2006. XRCC3 and


McCoull KD, Rindgen D, Blair IA, Penning TM. 1999. Synthesis and characterization of


Penning TM, Ohnishi ST, Ohnishi T, Harvey RG. 1996. Generation of reactive oxygen species during the enzymatic oxidation of polycyclic aromatic hydrocarbon trans-dihydrodiols catalyzed


Plante BJ, Cooper GS, Baird DD, Steiner AZ. 2010. The impact of smoking on antimullerian
hormone levels in women aged 38 to 50 years. Menopause 17(3):571-576.


methyl-6-phenylimidazo(4,5-b)pyridine (PhIP) in human urine after consumption of charbroiled or fried beef. Mutat Res 506-507:163-173.


Sun JD, Wolff RK, Kanapilly GM. 1982. Deposition, retention, and biological fate of inhaled benzo[a]pyrene adsorbed onto ultrafine particles and as a pure aerosol. Toxicol Appl Pharmacol 65(2):231-244.


Trapp C, McCullough AK, Epe B. 2007. The basal levels of 8-oxoG and other oxidative modifications in intact mitochondrial DNA are low even in repair-deficient (Ogg1(-/-)/Csb(-/-)) mice. Mutat Res 625(1-2):155-163.


239


